



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Biológicas

Escuela Profesional de Ciencias Biológicas

**Caracterización genética de poblaciones humanas de
tres regiones del Perú, para su identificación mediante
el uso de 30 marcadores de inserción y delección**

TESIS

Para optar el Título Profesional de Biólogo, Genetista,
Biotecnólogo

AUTOR

Luis Armando ZUÑIGA CCOICCA

ASESOR

Rina Lastenia RAMÍREZ MESÍAS

Lima, Perú

2018



Universidad Nacional Mayor De San Marcos
(Universidad del Perú, Decana de América)

FACULTAD DE CIENCIAS BIOLÓGICAS

**ACTA DE SESIÓN PARA OPTAR AL TÍTULO PROFESIONAL DE
BIÓLOGO GENETISTA BIOTECNÓLOGO
(MODALIDAD: SUSTENTACIÓN DE TESIS)**

Siendo las 16:20 horas del 31 de octubre de 2018, en el Salón de Grados de la Facultad de Ciencias Biológicas y en presencia del jurado formado por los profesores que suscriben, se dio inicio a la sesión para optar al **Título Profesional de: Biólogo Genetista Biotecnólogo** de **LUIS ARMANDO ZUÑIGA CCOICCA**.

Luego de dar lectura y conformidad al expediente N° 014-EPGB-2018, el titulado expuso su tesis: **"CARACTERIZACIÓN GENÉTICA DE POBLACIONES HUMANAS DE TRES REGIONES DEL PERÚ, PARA SU IDENTIFICACIÓN MEDIANTE EL USO DE 30 MARCADORES DE INSERCIÓN Y DELECIÓN"** y el Jurado efectuó las preguntas del caso calificando la exposición con la nota 19.....calificativo: Aprobado con máximos honores

Finalmente, el expediente será enviado a la Escuela Profesional de Genética y Biotecnología, y al Consejo de Facultad para que se apruebe otorgar el **Título Profesional de Biólogo Genetista Biotecnólogo** a **LUIS ARMANDO ZUÑIGA CCOICCA** y se eleve lo actuado al Rectorado para conferir el respectivo título, conforme a ley.

Siendo las 18:00 horas se levantó la sesión.

Ciudad Universitaria, 31 de octubre de 2018.

Dra. MONICA ARAKAKI MAKISHI
(PRESIDENTA)

Dra. RINA RAMÍREZ MESIAS
(ASESORA)

Mg. FERNANDO RETUERTO PRIETO
(MIEMBRO)

Blga. NILDA OLIVEROS RODRIGUEZ
(MIEMBRO)

DEDICATORIA

A mis padres Modesto Zuñiga Ccorisapra y Olimpia Ccoicca Atao, quienes son para mí modelos de unión, perseverancia, confianza y humildad. A Jenny por su amor y apoyo en todo momento de nuestro desarrollo como padres, como profesionales y como personas. A Camila, mi hija, quien me inspira a seguir avanzando con mucha felicidad y actitud positiva pensando siempre en un futuro mejor para todos. A mis siete hermanos menores quienes con mucho optimismo y dedicación vienen formándose como personas íntegras de bien para nuestra sociedad.

AGRADECIMIENTOS

Escribir esta parte de la tesis no es tan fácil para mí, cuando pienso en todas las personas que contribuyeron en ella, que me apoyaron e hicieron posible sustentar este trabajo del cual aprendí muchísimo académicamente y como persona, me doy cuenta que se trata de importantes personas que siempre estuvieron apoyándome.

En primer lugar, quiero agradecer a mi orientadora la doctora Rina Lastenia Ramírez Mesías docente e investigadora de la Universidad Nacional Mayor de San Marcos, por brindarme sus enseñanzas y consejos no solo en la elaboración de la tesis, sino también en mi formación como persona, enseñándome a no desistir ante ningún problema, a asumir los retos de la vida con optimismo. La doctora Rina, a pesar de tantas actividades que lleva bajo su responsabilidad, en cada saludo siempre dispone de un espacio para dialogar y escuchar a sus tesisas, para hacer preguntas y aconsejar, siempre pensando en el presente y en el futuro de los jóvenes. Por todo ello estoy muy agradecido y tengo el compromiso personal de compartir lo aprendido con otros jóvenes ahora y más adelante.

Asimismo, expreso mis agradecimientos al Ministerio Público Fiscalía de la Nación. Primero por abrirme las puertas y permitirme ser parte de sus valiosos profesionales, en segundo lugar por brindarme las condiciones formales autorizándome desarrollar la tesis en el Laboratorio de Biología Molecular y Genética. Agradecer también a la Jefatura Nacional del Instituto de Medicina Legal y Ciencias Forenses, en ella al doctor Segundo Eleazar Aliaga Viera quien a pesar de sus múltiples responsabilidades como Jefe Nacional del Instituto de Medicina Legal y Ciencias Forenses siempre expresó directamente su preocupación por la sustentación de esta tesis y su valor en mi formación como profesional.

De la misma manera, a la Escuela del Ministerio Público Fiscalía de la Nación, en ella al doctor Jaime Javier Villanueva Barreto por su aprobación a mi propuesta de investigación sin la cual no hubiera sido posible desarrollarla formalmente; también mi agradecimiento a Maria Estilita Bermejo Valladolid Sub Gerente del Laboratorio de Biología Molecular y Genética del Instituto de Medicina Legal y Ciencias Forenses del Ministerio Público por su apoyo en las gestiones pertinentes para el desarrollo de este estudio , también en el mismo laboratorio agradezco a Edgardo Delgado Ramos quien de manera incondicional y con mucha predisposición pudo apoyarme con la obtención de los perfiles genéticos, el respectivo análisis de los electroferogramas y obtención de los genotipos.

Muchos de los conceptos que pude comprender y discutir al terminar esta tesis los debo a las enseñanzas de una maravillosa persona, la doctora Leonor Gusmão, a quien tuve el honor de conocer en el año 2017 en el Laboratorio de Diagnóstico por DNA de la Universidad del Estado de Rio de Janeiro (UERJ) en Brasil. Expreso mi agradecimiento a la doctora Leonor por su inmensa paciencia, por su modo tan práctico y sencillo de explicar temas complejos haciéndolos fácil de comprender, por la entrega que pone en todo lo que hace, por difundir el inmenso respeto por el prójimo, por su humildad y por estar siempre dispuesta a apoyar a todo aquel que desea investigar temas de genética de poblaciones humanas, bases de datos de ADN, y otros temas afines a la genética forense.

Del mismo modo, agradezco a Rui Pereira, del Instituto de Patología e Inmunología Molecular de la Universidad de Porto (IPATIMUP) Portugal. Rui a pesar de sus múltiples actividades siempre pudo responder a mis consultas, ya sea por e-mails, o por otros medios de comunicación propios de estos tiempos como: Whatsapp o Skype. Su experiencia con los marcadores INDELs y su rigurosidad con los análisis de estos marcadores han sido de mucha ayuda en mi formación, aprendí bastante gracias a sus enseñanzas.

También expreso mi agradecimiento al doctor Oscar García de la Unidad de Ciencias Forenses, Sección de Genética Forense de la Policía del País Vasco, quien amablemente contribuyó a este trabajo con material suplementario de genotipos INDELs del País Vasco y España Central de la publicación “Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations”. De la misma manera mi agradecimiento a la doctora Maria Saiz Guinaldo del Laboratorio de Identificación Genética, Departamento de Medicina Legal, Toxicología y Antropología Física de la Universidad de Granada-España, quien amablemente facilitó material suplementario de genotipos INDELs de Uruguay de la publicación “Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population”. Estos datos permitieron realizar análisis comparativos de distancias genéticas (F_{st}), análisis de Varianza Molecular (AMOVA) y estructura genético-poblacional a nivel intercontinental, lo cual fue posible realizar gracias al apoyo de los doctores Oscar García y Maria Saiz Guinaldo.

Finalmente gracias a toda mi familia; a mis padres Modesto Zuñiga Ccorisapra y Olimpia Ccoicca Atao, quienes con su ejemplo me enseñaron a no rendirme ante ningún reto de la vida, a actuar con humildad y respeto y siempre con optimismo, buen humor y sobre todo en unión como familia. Del mismo modo, agradecer infinitamente a Jenny y Camila quienes alimentan todos los días estos deseos de superación en mí como padre, como esposo y como profesional. A mis siete hermanos menores quienes también motivan mi desarrollo como referente para ellos, como hermano mayor, pensando en ellos siempre busco actuar con justicia e igualdad. A todos los mencionados, muchas gracias.

Este trabajo se realizó con la participación de las siguientes instituciones:

Escuela Profesional de Genética y Biotecnología. Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas.



UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA

Laboratorio de Biología Molecular y Genética, Instituto de Medicina Legal y Ciencias Forenses. Ministerio Público, Fiscalía de la Nación, Escuela del Ministerio Público.



MINISTERIO PÚBLICO
FISCALÍA DE LA NACIÓN

Laboratório de Diagnóstico por DNA- UERJ. Universidade do Estado do Rio de Janeiro- Brasil.



**UNIVERSIDADE DO ESTADO DO
RIO DE JANEIRO**

ÍNDICE GENERAL

DEDICATORIA	I
AGRADECIMIENTOS	II
RESUMEN	III
ABSTRACT	IV
1. INTRODUCCIÓN	1
2. HIPÓTESIS	3
3. OBJETIVOS	3
4. MARCO TEÓRICO	4
4.1. Polimorfismos de inserción y delección (DIPs o INDELs)	5
4.2. Distribución cromosómica de INDELs en humanos	5
4.3. Distribución de INDELs relativo a los genes humanos	6
4.4. Los INDELs y las enfermedades en humanos	6
4.5. Los INDELs como marcadores genéticos en ciencias forenses	6
4.6. Relevancia de los INDELs como marcadores genéticos en ciencias forenses.....	7
5. MATERIALES Y MÉTODOS	8
5.1. Marcadores de inserción y delección INDELs	8
5.2. Población de estudio	10
5.2.1 Criterios de inclusión de la muestra	11
5.2.2 Criterios de Exclusión de la Muestra	11
5.2.3 Procedimiento de selección de las muestras	11
5.3. Extracción de ADN	11
5.4 Amplificación por PCR múltiplex	12
5.4.1. Condiciones de la PCR	13
5.4.2 Configuración de termociclado	13
5.5. Electroforesis capilar	13
5.5.1 Materiales empleados en analizador genético	14
5.5.2 Preparación de la muestra para electroforesis capilar	15
5.5.3. Obtención de datos de electroforesis capilar	19

5.6. Análisis de datos	19
5.7. Análisis genético poblacional	21
5.7.1. Análisis de estructura poblacional	21
5.7.2. Análisis de estructura poblacional comparado con datos de poblaciones europeas y sudamericanas	22
5.7.3. Prueba de Equilibrio de Hardy–Weinberg (EHW)	23
5.7.4. Prueba F_{st}	24
5.8. Índices de diversidad estándar	26
5.8.1. Diversidad genética	26
5.8.2. Cálculo de heterocigosidad esperada y heterocigosidad observada	26
5.8.3. Estimación de frecuencias alélicas	27
5.8.4. Análisis de varianza molecular (AMOVA)	27
5.8.5. Estimación de desequilibrio de ligamiento	28
5.9. PARÁMETROS ESTADÍSTICOS DE INTERÉS FORENSE	28
5.9.1. Probabilidad de match (MP)	28
5.9.2. Poder de discriminación (PD)	29
5.9.3. Contenido de información polimórfica (PIC)	29
5.9.4. Poder de exclusión (PE)	29
5.9.5. Índice típico de paternidad (TPI)	30
6. ANÁLISIS DE ESTRUCTURA POBLACIONAL	30
7. RESULTADOS	31
7.1. Análisis de estructura poblacional en tres regiones del Perú	31
7.2. Análisis de estructura poblacional comparado con datos de poblaciones europeas y sudamericanas	32
7.3. Prueba de equilibrio de Hardy–Weinberg	34
7.4. Análisis de varianza molecular (AMOVA)	38
7.5. Análisis de diferenciación poblacional F_{st}	39
7.6. Estimación del desequilibrio de ligamiento	40
7.7. Inferencia de historia evolutiva	41

8. ANÁLISIS DE PARAMETROS ESTADÍSTICOS DE INTERÉS FORENSE	42
9. DISCUSIÓN	44
10. CONCLUSIONES	46
11. RECOMENDACIONES	47
12. REFERENCIAS BIBLIOGRÁFICAS	49

ÍNDICE DE FIGURAS

Figura 1 Distribución de Indels en el genoma humano	5
Figura 2 Electroferograma o perfil genético obtenido en equipo ABI3500 de ochocapilares.....	14
Figura 3. Analizador genético modelo ABI 3500 de 8 capilares (manual de Applied Biosystems® 3500/3500xL Genetic Analyzer	14
Figura 4. Posición adecuada de componentes líquidos inyectados	17
Figura 5. Electroferograma del estándar de tamaño de 550 (BTO), con sus respectivos fragmentos de longitud en pares de bases	18
Figura 6. Electroferograma de una muestra obtenido en Analizador Genético ABI3500 de Applied Biosystems mediante programa Gene Mapper ID-X	20
Figura 7. Datos recodificados de perfiles genéticos INDELs en formato Excel	21
Figura 8. Mapa del Perú con las tres regiones estudiadas. En rojo la región Norte, en verde la región Centro, en azul la región Sur	22
Figura 9. Mapa de América del Sur (Perú y Uruguay) y una parte del continente europeo (España Central y País Vasco), con los cuales se hizo el análisis de estructura poblacional a escala intercontinental	23
Figura 10. Resultados del análisis de estructura poblacional de las tres regiones estudiadas del Perú	31
Figura 11. Resultados del Análisis de estructura poblacional a nivel Intercontinental	33
Figura 12. Arbol de distancias genéticas de la población Peruana	41

Figura 13. Árbol de distancias Genéticas de la población peruana y España Central, País Vasco y Uruguay	42
--	----

ÍNDICE DE TABLAS

Tabla 1. Resumen de participantes de la región norte	10
Tabla 2. Resumen de participantes de la región centro	10
Tabla 3. Resumen de participantes de la región sur	10
Tabla 4. Condiciones de PCR	13
Tabla 5. Configuración de termociclado.* Temperatura de activación de la DNA polimerasa (Hot-Start)	13
Tabla 6. Consumibles propios del equipo de electroforesis capilar ABI 3500	15
Tabla 7. Componentes y volúmenes de reactivos para electroforesis capilar	17
Tabla 8. Número de agrupamientos con mayor verosimilitud ensayado con 136 genotipos	32
Tabla 9. Número de agrupamientos con mayor verosimilitud ensayado a nivel Intercontinental	33
Tabla 10. Resultados del análisis de equilibrio de Hardy Weinberg	34
Tabla 11. Comparación del equilibrio de Hardy-Weingberg en cuatro poblaciones mediante 30 INDELs	36
Tabla 12. Comparación de la diversidad genética con otras poblaciones	37
Tabla 13. Resultados del análisis de varianza molecular (AMOVA) entre Norte, Centro y Sur. Obtenido mediante el programa Arlequin V.3.5	38
Tabla 14. Significancia del análisis de varianza molecular	39
Tabla 15. Tabla de significancia del desequilibrio de ligamiento (nivel de significancia 0.05)	40

ÍNDICE DE CUADROS

Cuadro 1. Especificaciones de los 30 marcadores INDELs. Se muestra: Localización cromosómica,código de acceso GenBank, motivo y alelo referente	9
Cuadro 2. Marcadores fluorescentes de BT5	15
Cuadro 3. Matriz estándar y correlación con cada locus y códigos asignados	16
Cuadro 4. Parámetros configurados para ejecutar electroforesis capilar de 30 marcadores (DIPlex Kit) en analizador genético ABI 3500 de Applied Biosystems	18
Cuadro 5. Correlación de la significancia (P-value) con las distancias genéticas (Fst) entre las poblaciones de Perú, España, País Vasco y Uruguay	39
Cuadro 6. Comparación de valores de Fst y niveles de significancia entre las poblaciones de Perú, España Central, País Vasco y Uruguay	40
Cuadro 7. Parámetros estadísticos de interés forense: Parámetros de paternidad, contenido de Información Polimórfica (PIC), Probabilidad de Match (PM) y Poder de Discriminación (PD)	43

RESUMEN

En este estudio se caracterizó genéticamente 136 peruanos mestizos de las regiones norte, centro y sur del Perú, empleando un conjunto de 30 marcadores autosómicos no codificantes de inserción y delección denominados INDELs. Se evaluó la diversidad genética como una medida de la heterosigosidad media en la población y se obtuvieron las frecuencias alélicas y genotípicas para los 30 marcadores estudiados. De los análisis de varianza molecular (AMOVA) y distancias genéticas entre poblaciones (F_{st}) no se observó diferencias significativas entre los grupos norte, centro y sur del Perú, pero si una alta diversidad genética a nivel de individuos (97.86%), es decir que la población mestiza del Perú se presenta en estos análisis como una unidad poblacional sin subestructura genética. Ninguno de los 30 marcadores INDELs mostraron desviaciones al Equilibrio de Hardy–Weinberg.

Por otro lado los análisis de estructura poblacional en el Perú no muestran algún grado de agrupamiento que evidencie subestructura, sin embargo a una escala intercontinental se realizó una comparación con datos que emplearon el mismo sistema de marcadores en la población de España, País Vasco y Uruguay resultando un valor de K igual a 02 (dos), donde se observa que el Perú se muestra como un grupo poblacional con muy poco componente genético español a diferencia de Uruguay donde el componente español es predominante. De estos resultados se infiere que la población peruana posee mayormente un componente genético americano nativo que podría ser andino o amazónico.

Adicionalmente se evaluó los parámetros estadísticos de interés forense. Este sistema de 30 marcadores muestra un alto Poder de Discriminación (99.99999999%) y una baja probabilidad de coincidencia (0.0000000001426561820), estas variables son importantes en genética forense ya que posee una alta capacidad de discriminar un perfil genético de ADN de otro que se haya encontrado en la misma escena del crimen, adicionalmente al poseer una baja probabilidad de coincidencia significa que la probabilidad de que dos individuos no relacionados tengan el mismo perfil genético en la escena del crimen es prácticamente cero.

PALABRAS CLAVE: INDELs autosómicos, peruanos mestizos, estructura genética, diversidad genética, Poder de Discriminación (PD), Probabilidad de coincidencia (PM), Genética Forense.

ABSTRACT

In this study were characterized genetically 136 Peruvian mestizos from the north, center and south regions of Perú, using a set of 30 non-coding autosomal insertion and deletion markers called INDELs. Genetic diversity was evaluated as a measure of the mean heterosigosity in the population; the database of allelic and genotypic frequencies was obtained for the 30 markers studied. The Analysis of Molecular Variance (AMOVA) and genetic distances between populations (F_{st}) didn't showed significant differences between the northern, Central and Southern population, but high genetic diversity was observed at the individual level (97.86%), that is, the mestizo population of Peru is presented in this analisis as a unique population without genetic substructure. The 30 INDELs markers showed no deviations to the Hardy-Weinberg Equilibrium.

On the other hand, population structure analyzes in Peru do not show any degree of clustering which evidences substructure, however a intercontinental scale comparison was made with data that used the same marker system on the population of Spain, the Basque Country and Uruguay, resulting in a value of K equal to 02 (two), where Perú is presented as a population group with very little Spanish genetic component unlike Uruguay where the Spanish component is predominant. From these results, it is inferred in the study that the Peruvian population owns mainly a native American genetic component that could be Andean or Amazonian.

Additionally, the statistical parameters of forensic interest were analized. This system of 30 markers shows a high Discrimination Power (99.99999999%) and a low Matching Probability (0.0000000001426561820), these variables are very important in forensic genetics since it has a high capacity to discriminate a DNA genetic profile of another that has been found at the same crime scene. Having a low probability of coincidence, the probability that two unrelated individuals have the same DNA genetic profile at the crime scene is virtually zero.

KEY WORDS: Autosomic INDELS, Peruvian mestizos, genetic structure, genetic diversity, Power of Discrimination (PD), Probability of coincidence (PM), Forensic Genetics.

1. INTRODUCCIÓN

Actualmente los descubrimientos biotecnológicos y la implementación de instrumentos tecnológicos se vienen desarrollando de manera acelerada en todos los campos de la ciencia, es así que en el campo de la genética forense se ha desarrollado una gran variedad de metodologías basadas en la tecnología del ADN, la cual se emplea como herramienta para resolver casos de criminalística, identificación de personas, paternidad, identificación de desaparecidos en desastres masivos, en casos de delitos de lesa humanidad, desapariciones forzadas, asaltos sexuales y otros casos relacionados.

En todos los casos antes citados los resultados de exámenes de ADN y sus consecuentes conclusiones tienen un impacto directo sobre la aplicación de justicia y, por ende, supone en los científicos forenses una obligación y compromiso de proveer resultados, científicamente probables, acreditables mediante estándares internacionales. Asimismo, en concordancia con la constante implementación de nuevos instrumentos o sistemas de caracterización de ADN, los métodos que han de emplear implica que se encuentren caracterizados y estandarizados para el grupo poblacional dentro del cual se viene aplicando determinada metodología.

Partimos del principio de que las pruebas de ADN se enfocan ampliamente en el uso de polimorfismos genéticos para la identificación de individuos, empleando así diferentes tipos de marcadores genéticos tales como la amplificación aleatoria de ADN polimórfico (RAPDs), la región hipervariable del ADN mitocondrial (HV1, HV2), las regiones de secuencias simples repetitivas (ISSRs), las secuencias cortas repetitivas en tándem (STRs), los polimorfismos nucleotídicos únicos (SNPs), etc. De estos, en la actualidad los marcadores más utilizados a nivel mundial en genética forense son los STRs debido a su alto polimorfismo, su alta tasa de mutación, su poder de discriminación, su practicidad, facilidad de análisis, y por ser un método relativamente económico (Butler, 2005).

Sin embargo, existen algunas desventajas en los STRs. Primero, su poder de discriminación en casos de paternidad es limitado, en segundo lugar, los STRs exhiben una significativa tasa de mutación (hasta 0.05) que puede causar serios problemas en casos de paternidad donde no se tiene información de la madre o cuando un pariente del presunto padre tiene que ser considerado como padre alternativo (Wenk, 2004), también cuando se tienen todos los alelos correctamente genotipificados pero se encuentran mutaciones únicas ya sea en el presunto padre, la madre o el hijo en cuestión. Adicionalmente, debemos mencionar que los STRs tienen la limitación de no amplificar por completo sus marcadores por causa de alta degradación del material genético, ya sea por el proceso de descomposición o

por su antigüedad por lo que muchas veces no se obtienen perfiles genéticos completos, ya que se requieren largas hebras de ADN molde (Pereira & Gusmão, 2012).

Actualmente los científicos forenses han centrado su atención en otro tipo de variación genética que puede ser usada en el análisis de muestras degradadas y casos de paternidades con aparición de mutaciones únicas, son los llamados polimorfismos de inserción y deleción INDELs (Martin, 2012). Estos marcadores tienen la ventaja de presentar una baja tasa de mutación comparado con los STRs, presentan estabilidad genética y tamaño corto de amplicón (Zaumsegel et al., 2012) lo que los hace adecuados para obtener y analizar perfiles genéticos de muestras antiguas o degradadas (Hollard et al., 2011).

En el presente trabajo de investigación, realizamos un estudio de la estructura genética, su variabilidad y parámetros estadísticos forenses, como la probabilidad de match (MP), poder de discriminación (PD) contenido de información polimórfica (PIC), poder de exclusión (PE) y otros parámetros genético poblacionales. Para ello se tuvo la participación de 136 personas de diferentes ciudades del Perú quienes fueron agrupados según su procedencia en tres regiones geográficas; Norte, Centro y Sur. El sistema de marcadores genéticos empleados fue de 30 marcadores de Inserción y deleción INDELs (Investigator DIPLEX-Kit, Germany).

El conocimiento de los parámetros genético poblacionales antes citados permitirá poseer referencias o antecedentes de mucha importancia científica para el uso forense del sistema de INDELs en el Perú, en el análisis de muestras antiguas como restos óseos, muestras degradadas y en casos de paternidad con presencia de mutación. Asimismo, permitirá conocer las características genético poblacionales del Perú para este tipo de sistema. De esta manera se espera aportar conocimientos que contribuyan en la mejora continua de la calidad del trabajo y la investigación forense en el Perú.

2. HIPÓTESIS

2.1 Hipótesis Alternativa

Las tres regiones de poblaciones humanas analizadas presentan una estructura genética diferenciada empleando 30 marcadores de inserción y deleción INDELs y a su vez este sistema resulta eficiente en el análisis genético forense de cada grupo poblacional.

2.2 Hipótesis nula

Las tres regiones de poblaciones humanas analizadas no presentan una estructura genética diferenciada empleando 30 marcadores de inserción y deleción INDELs e independientemente este sistema resulta eficiente en el análisis genético forense de la población mestiza del Perú.

3. OBJETIVOS

3.1 Objetivo General

Caracterizar genéticamente poblaciones humanas de tres regiones del Perú, mediante 30 marcadores de inserción y deleción INDELs y analizar las variables genético poblaciones y parámetros estadísticos de interés forense empleados en identificación humana.

3.2 Objetivos Específicos

- Obtener perfiles genéticos de 136 peruanos mestizos procedentes del Norte, Centro y Sur del Perú empleando un set estándar de 30 marcadores de Inserción y Deleción (INDELs).
- Analizar la variabilidad intra e interpoblacional y su correlación entre el Norte, Centro y Sur del Perú.
- Evaluar la estructura genética de la población mestiza del Perú.
- Analizar los parámetros estadísticos de interés forense del set de 30 marcadores INDELs en la población peruana.
- Obtener las bases de datos de frecuencias alélicas y genotípicas de la población peruana.

4. MARCO TEÓRICO

Los primeros estudios sobre las causas de la variabilidad genética donde se reconoce a la replicación cromosómica como una condición mutacional que forma parte de la evolución, datan de hace más de 70 años (Bridges, 1936). Por otro lado, el mecanismo molecular de la duplicación cromosómica, la aparición de inserciones, deleciones, mutaciones puntuales y todos los demás tipos conocidos de polimorfismos, así como su frecuencia y sus implicancias en la evolución del genoma han sido esclarecidos recientemente en cromosomas de bacteria, mosca de la fruta y humanos (Lupski et al., 1996). Fue así que en el año 2001 ya se había observado que las inserciones y deleciones (Indels) son mutaciones abundantes en diversos modelos animales. Por ejemplo, estudios de variación genética en *Drosophila melanogaster* y *Caenorhabditis elegans* ya habían demostrado que los Indels representan entre 16 al 25% del total de polimorfismos genómicos en estas especies (Berger et al., 2001; Wicks et al., 2001). También en el 2001, otro estudio de variación genética en el cromosoma 22 humano reveló que los Indels representan el 18% del total de polimorfismos en este cromosoma (Dawson et al., 2001). Basados en estos estudios, Mills y colaboradores (2006) estimaron que los Indels probablemente representen entre el 16 al 25% del total de polimorfismos del genoma humano.

Asimismo, estudios realizados en humanos mediante Indels contribuyeron fundamentalmente al conocimiento de polimorfismos basados en inserciones y deleciones (Weber et al., 2002), ya que hasta entonces las investigaciones de las variantes del ADN en genética forense estaban enfocados al empleo de secuencias cortas repetitivas en tándem (STRs), polimorfismos nucleotídicos únicos (SNPs), ADN mitocondrial y eventualmente el uso de Indels. Hasta que Weber y colaboradores (2002) identificaron y caracterizaron 2000 Indels humanos distribuidos a lo largo de nuestro genoma con diferentes longitudes y frecuencias alélicas, aportando significativamente a este conocimiento. Posteriormente, Mills y colaboradores (2006) publican un mapa inicial de Indels del genoma humano, en el que reportan 415,436 polimorfismos únicos de Indels que fueron identificados mediante secuenciamiento y análisis computacional. De esta manera se supo que las longitudes de los Indels varían en un rango de 1 a 9,989 pares de bases y presentan una longitud casi igual a las inserciones y deleciones de chimpancés (Mills et al., 2006). Asimismo, clasificaron los Indels en cinco grandes grupos como se muestra: (1) Deleciones e inserciones de nucleótido único, (2) expansiones monoméricas de pares de bases, (3) expansión de 2 a 15 pares de bases como unidades repetitivas, (4) inserción de transposones y (5) Indels que contienen secuencias aleatorias variables (Mills et al., 2006).

4.1 Polimorfismos de Inserción y Delección (DIPs o INDELs)

Los indels son polimorfismos de longitud que se generan por inserción o delección de uno o más nucleótidos en el genoma (Pereira et al., 2009). Existen más de 415 000 polimorfismos únicos de Indels distribuidos a lo largo del genoma humano con una densidad promedio de un Indel cada 7.2 kb de ADN (Mills et al., 2006). Asimismo, las inserciones y delecciones en nuestro genoma están distribuidas casi equitativamente: el 47% del total de Indels corresponde a inserciones en el genoma mientras que el 53% corresponde a delecciones, estas observaciones realizadas por Mills y colaboradores en el año 2006 son similares a las proporciones de distribución de Indels en el genoma de Chimpancés observado por los mismos investigadores, previo al análisis de Indels en el genoma humano.

4.2 Distribución Cromosómica de INDELs en Humanos

La distribución de INDELs tiene correlación directa con la cantidad de ADN presente en los cromosomas humanos (Fig.1), sin embargo en algunos casos (cromosomas 4, 5, 8, 14,18 y 20) presentan un incremento en la cantidad de INDELs, a diferencia de los otros cromosomas autosómicos, mientras que los cromosomas X e Y aparentemente presentan menor variación de INDELs en contraste con los cromosomas autosómicos, Sin embargo, tanto los cromosomas sexuales como los cromosomas autosómicos presentan en promedio una distribución de la variación de INDELs, de un INDEL por cada 5.1 a 13.2 Kb de ADN genómico. Estos resultados fueron obtenidos por Mills y colaboradores (2006) en un estudio de mapeo de INDELs en el genoma humano

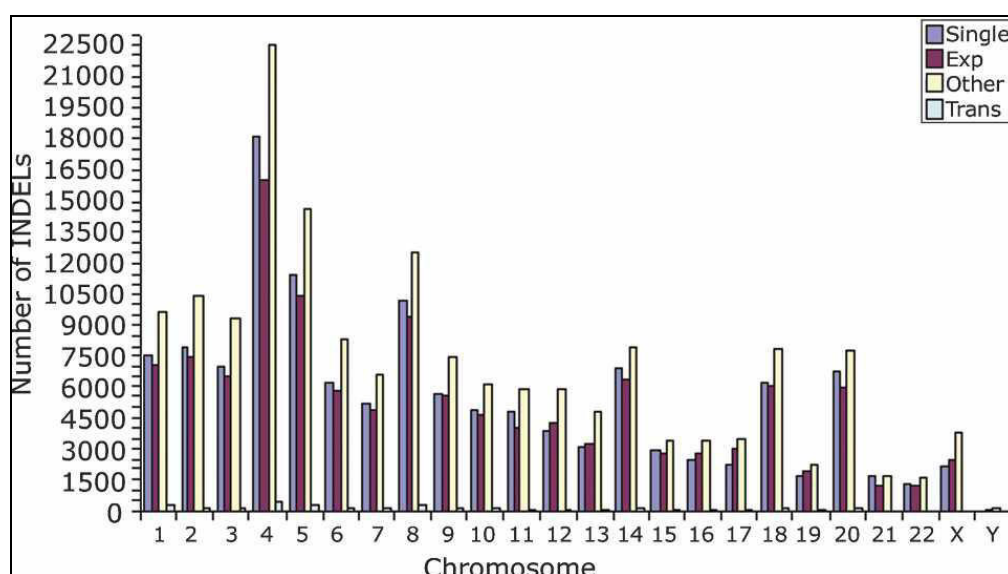


Figura 1. Distribución de Indels en el genoma humano (Mills et al., 2006).

4.3 Distribución de INDELs relativo a los genes humanos

Del total de INDELs existentes en el genoma humano el 35.7% se encuentran localizados dentro de genes conocidos a nivel de los promotores y exones donde se estima que podrían estar influenciando fuertemente la función de los genes donde se encuentran, así por ejemplo estarían causando errores en los marcos de lectura que conllevan a la terminación prematura de las proteínas que codifican. Muchos de los INDELs podrían estar bloqueando o disminuyendo la función de los genes donde se encuentran por lo que podrían constituir alelos recesivos en la población humana (Mills et al., 2006). En ese sentido se infiere que los INDELs tienen un impacto sobre la función de los genes humanos.

4.4 Los INDELs y las enfermedades en humanos

Los polimorfismos de INDELs son de gran importancia en la salud humana porque pueden alterar los fenotipos y causar algunas enfermedades. Una de las enfermedades más comunes en humanos es la Fibrosis Cística que frecuentemente es causado por un alelo en el gen CFTR que contiene una delección de tres pares de bases, esta delección conlleva a la eliminación de un aminoácido que constituye la proteína, desencadenando así la enfermedad de fibrosis cística (Collins et al., 1987).

Las enfermedades por expansión de trinucleótidos tales como el Síndrome del X Frágil son causadas por inserciones de ADN resultado de la expansión de unidades de trinucleótidos repetitivos cortos (Warren et al., 1987). Los elementos transponibles al integrarse dentro de los genomas producen inserciones de ADN que pueden causar enfermedades en humanos tales como hemofilia, neurofibromatosis, distrofia muscular y cáncer (Ostertag & Kazazian, 2001).

4.5 Los INDELs como marcadores genéticos en ciencias forenses

Si bien es cierto que la genética forense está basada en el empleo de polimorfismos de ADN, sin embargo, al respecto se conoce poco acerca de los tipos y número de variantes dentro de la población humana (Weber et al., 2002). La mayoría de los polimorfismos de ADN humano están divididos en dos grupos: Aquellos basados en sustituciones nucleotídicas (comúnmente denominados SNPs) y aquellos basados en la inserción o delección de uno o más nucleótidos (INDELs). Los INDELs que bien podrían tener alelos múltiples (multialélicos) o solamente dos alelos (dialélicos) y la mayoría de los INDELs multialélicos están basados en repeticiones nucleotídicas en tándem conocidos como STRs o “Microsatélites”, son éstos los polimorfismos de ADN predominantemente utilizados en estudios de genética

humana desde los 1990s (Weber et al., 2002). Y más recientemente millones de SNPs candidatos para estudios de genética humana han sido identificados (Sachidanandam et al., 2001)

Los INDELs dialélicos son los que menos atención recibieron para estudios de genética de poblaciones hasta hace algunos años (Weber et al., 2002), éstos presentan diferencias significativas en las longitudes de sus alelos, en raros casos la diferencia de longitud de sus alelos puede ser de decenas o de cientos de pares de kilobases (Lupski et al., 1996), algunos INDELs dialélicos se diferencian por la inserción de un retrotransposon, tales como un elemento *Alu* o *L1* (Watkins et al., 2001).

Weber y colaboradores (2002) reportan la identificación y caracterización de 2000 polimorfismos de inserción y delección en el genoma humano, los cuales fueron identificados por alineamiento múltiple de secuencias de ADN genómico y/o cDNAs, obtuvieron así las frecuencias alélicas de INDELs, así como la proporción de mutaciones de inserción y delección, la cual fue de 4 a 1 respectivamente, realizándose dichas medidas en poblaciones de Europa, África, Japón, y Nativos Americanos. Asimismo, precisaron que los alelos nuevos encontrados presentaron una baja frecuencia, comparado con la frecuencia de los alelos antiguos, y que esta tendencia fue más pronunciada en la población africana quienes probablemente están más relacionados con los cuatro grupos poblacionales humanos modernos, asimismo concluyeron que los INDELs dialélicos comprenden el 8% del total de polimorfismos humanos y que su abundancia y facilidad de análisis los hacen útiles para muchas aplicaciones.

Pereira y colaboradores (2009) genotipificaron 38 INDELs dialélicos autosómicos no codificantes cuyas longitudes fueron inferiores a 160 pares de bases que previamente habían sido reportados como polimórficos en poblaciones de África, Europa y Asia mediante PCR multiplex y electroforesis capilar. En dicho estudio se reporta la obtención de perfiles genéticos completos usando 0.3 ng de ADN en muestras degradadas en los cuales no se había obtenido perfiles genéticos completos empleando STRs, proponiendo así el empleo potencial de INDELs en estudios de identificación humana como una alternativa entre los SNPs y STRs.

4.6 Relevancia de los INDELs como marcadores genéticos en ciencias forenses

En muchos casos, según el estado de degradación de las muestras, resulta necesario emplear otros métodos de amplificación de ADN que consten de pequeños tamaños de amplicones tales como los SNPs, miniSTRs e INDELs (Weber et al., 2002; Mills et al., 2006). Recientemente la comunidad forense viene enfocándose en el uso de INDELs como alternativa forense, debido a

determinadas ventajas que estos marcadores presentan, tales como: 1. Baja tasa de mutación comparado con los STRs lo cual los hace apropiados para análisis de parentesco cuando ocurren mutaciones en loci de STRs (Pereira et al., 2009), 2. Están ampliamente distribuidos a lo largo del genoma y pueden amplificarse por PCR multiplex en pequeños fragmentos lo cual los hace adecuados para el análisis de muestras degradadas o muestras antiguas (Hollard et al., 2011; Romanini et al., 2012), 3. No presentan Stutters o microvariantes en el análisis de electroferogramas lo que los hace apropiados en la interpretación de mezclas (Carvahlo et al., 2011), 4. También son considerados buenos candidatos como marcadores informativos de ancestría (AIMs) ya que permiten identificar subestructura genética en poblaciones mezcladas (Santos et al., 2010; Pereira et al., 2012). Cabe resaltar como se precisó en una de sus ventajas la relevancia de los INDELs en investigaciones de paternidad, cuando al ser analizados con STRs aparecen mutaciones en el presunto padre constituyendo incompatibilidades hereditarias mendelianas, esto debido a la alta tasa de mutación de los STRs, en dichos casos resulta importante el empleo de marcadores con baja tasa de mutación como los SNPs o los INDELs (Pereira et al., 2009). Finalmente otra ventaja de los INDELs es que pueden ser analizados fácilmente mediante procesos de automatización que actualmente se utilizan en laboratorios forenses.

5. MATERIALES Y MÉTODOS

5.1. Marcadores de Inserción y Delección INDELs

En el presente estudio se utilizó 30 marcadores dialélicos de Inserción y Delección (DIPs o INDELs) más un marcador de amelogenina para determinar sexo. Estos marcadores pueden ser amplificados por PCR-múltiplex y constan de pequeños tamaños de amplicones (50-150 pb) (Pimenta et al., 2010; Li et al., 2011). Los 30 marcadores INDELs se encuentran distribuidos en 19 cromosomas autosómicos y han sido diseñados para uso forense por la empresa QIAGEN-Germany (Investigator-DIPplex Kit) para la obtención de perfiles genéticos de manchas sanguíneas, fluidos corporales, cabello y restos óseos de cadáveres antiguos y recientes. En el Cuadro1 se muestra las especificaciones moleculares sobre estos marcadores, tales como su ubicación cromosómica, el código de acceso GenBank, motivo y alelo referente.

Adicionalmente al uso forense que actualmente se le viene dando, estos marcadores también son adecuados para estudios de genética de poblaciones en humanos debido a su amplia distribución a lo largo del genoma; asimismo, han mostrado ser polimórficos con diferencias en las frecuencias alélicas en estudios de poblaciones (Comas et al., 2000; Weber et al., 2002).

Cuadro 1. Especificaciones de los 30 marcadores, se muestra: Localización cromosómica, código de acceso al GenBank, motivo y alelo referente (Investigator DIPplex Handbook. 2013)

N°	DIP-LOCUS	LOCALIZACIÓN CROMOSÓMICA	N° ACCESIÓN GEN BANK	MOTIVO (DIP)	ALELO REFERENTE	COLOR ELECTROFORESIS CAPILAR
	Amelogenin X	Xp22.1-22.3	M55418		X	DIPplex BLUE
	Amelogenin Y	Yp11.2	M55419		Y	
1	HLD77	7q31.1	rs1611048	TAAG	+DIP	
2	HLD45	2q31.1	rs2307959	CACG	–DIP	
3	HLD131	7q36.2	rs1611001	TGGGCTTATT	+DIP	
4	HLD70	6q16.1	rs2307652	AGCA	–DIP	
5	HLD6	16q13	rs1610905	GCAGGACTGG GCACC	–DIP	
6	HLD111	17p11.2	rs1305047	CACA	–DIP	
7	HLD58	5q14.1	rs1610937	AGGA	+DIP	
8	HLD56	4q25	rs2308292	TAAGT	+DIP	DIPplex GREEN
9	HLD118	20p11.1	rs16438	CCCCA	–DIP	
10	HLD92	11q22.2	rs17174476	GTTT	–DIP	
11	HLD93	12q22	rs2307570	ACTTT	–DIP	
12	HLD99	14q23.1	rs2308163	TGAT	–DIP	
13	HLD88	9q22.32	rs8190570	CCACAAAGA	+DIP	
14	HLD101	15q26.1	rs2307433	GTAG	–DIP	
15	HLD67	5q33.2	rs1305056	CTACTGAC	–DIP	
16	HLD83	8p22	rs2308072	AAGG	–DIP	DIPplex YELLOW
17	HLD114	17p13.3	rs2307581	TCCTATTCTACT CTGAAT	–DIP	
18	HLD48	2q11.2	rs28369942	GACTT	–DIP	
19	HLD124	22q12.3	rs6481	GTGGA	–DIP	
20	HLD122	21q22.11	rs8178524	GAAGTCTGAGG	–DIP	
21	HLD125	22q11.23	rs16388	ATTGCC	–DIP	
22	HLD64	5q12.3	rs1610935	GACAAA	+DIP	
23	HLD81	7q21.3	rs17879936	GTAAGCATTGT	–DIP	
24	HLD136	22q13.1	rs16363	TGTTT	–DIP	DIPplex RED
25	HLD133	3p22.1	rs2067235	CAACCTGGATT		
26	HLD97	13q12.3	rs17238892	AGAGAAAGCTG AAG	–DIP	
27	HLD40	1p32.3	rs2307956	GGGACAGGTGG CCACTAGGAGA	+DIP	
28	HLD128	1q31.3	rs2307924	ATTAAATA	–DIP	
29	HLD39	1p22.1	rs17878444	CCTAAACAAAAA TGGGAT	–DIP	
30	HLD84	8q24.12	rs3081400	CTTTC	–DIP	

HLD: Human Locus DIP; -DIP: Deletion; +DIP: Insertion, (Investigator-DIPplex-Handbook QIAGEN-Germany, 02/ 2013)

5.2. Población de estudio

Se tomaron de manera aleatoria 136 muestras sanguíneas en tarjetas FTA (Flinders Technology Associates) de personas no emparentadas procedentes de diferentes ciudades del Perú, de quienes previo consentimiento informado, se tomaron muestras sanguíneas durante el año 2008 en las diferentes sedes del Instituto de Medicina Legal a nivel nacional, las cuales se fueron acopiando en el Laboratorio de Biología Molecular y Genética de Lima. Para los análisis genéticos se agruparon las muestras en 3 regiones; Norte, Centro y Sur con el siguiente detalle:

Tabla 01. Resumen de participantes de la Región Norte

CIUDAD	N° INDIVIDUOS	REGIÓN
Cajamarca	12	NORTE
La Libertad -Lambayeque	14	
Amazonas -San Martin	12	
Piura –Tumbes	17	
TOTAL	55	

Tabla 2. Resumen de participantes de la Región Centro

CIUDAD	N° INDIVIDUOS	REGIÓN
Huánuco-Ucayali	17	CENTRO
Junin -Cerro de Pasco	12	
Lima-Ica	14	
Ancash-Norte de Lima	16	
TOTAL	59	

Tabla 3. Resumen de participantes de la Región Sur.

CIUDAD	N° INDIVIDUOS	REGIÓN
Puno-Arequipa	13	SUR
Ayacucho-Cusco	9	
TOTAL	22	

5.2.1 Criterios de inclusión de la muestra

- Que los individuos cuyas muestras se tomaron pertenecieran a la base de datos del Laboratorio de Biología Molecular y Genética del Instituto de Medicina Legal del Perú.
- Que el lugar de procedencia y residencia del individuo al que se tomó la muestra perteneciera a uno de los departamentos de las tres regiones (Norte, Centro y Sur) en las que se dividió el presente estudio.
- Que la muestra sea sanguínea y que se haya recolectado en tarjetas FTA para garantizar la conservación del material genético.

5.2.2 Criterios de exclusión de la muestra

- Que la muestra no cuente con consentimiento informado.
- Que la muestra tenga un soporte distinto a las tarjetas FTA (gasa, papel filtro u otros)

5.2.3 Procedimiento de selección de las muestras

- a. Revisión de los registros de casos que contengan muestras procedentes de las tres regiones en que se dividió el presente estudio (Norte, Centro, Sur)
- b. Revisión de los archivos físicos de los casos a fin de verificar procedencia y consentimiento informado.
- c. Asignación de códigos a las muestras seleccionadas y agrupamiento por departamentos y luego por región.
- d. Selección de 136 muestras sanguíneas.

5.3 Extracción de ADN

Se empleó micro-Puncher de 0.5 mm para obtener cortes circulares de 0.5 mm de diámetro (micro-punch) de las muestras sanguíneas en tarjetas FTA. Se obtuvo 8 cortes circulares por cada muestra (tarjeta FTA) las cuales fueron introducidas en tubos de 1.5 mL rotulados con los códigos asignados en el procedimiento de selección de muestra.

Las muestras sanguíneas fueron sometidas al siguiente procedimiento:

1. Cortar 8 fragmentos circulares de 0.5 mm de diámetro de tarjeta FTA con muestra de sangre e introducirlo en tubos de 1.5mL.
2. Al tubo de 1.5 mL que contiene la muestra de sangre añadir 300 uL de solución FTA, esperar 2 minutos y dar vortex, luego retirar y desechar dicha solución con una micropipeta.
3. Vuelve a añadir 200 uL de solución FTA y repetir el procedimiento anterior. A medida que se va vortexeando los micro-punchs se tornarán de color blanco y se irán eliminando los eritrocitos (líquido rojizo).
4. Añadir nuevamente 200 uL de solución FTA y repetir el procedimiento anterior.
5. Añadir 200 uL de buffer TE 10:1, esperar 2 minutos, luego dar vortex, retirar y eliminar el sobrenadante empleando una micropipeta.
6. Repetir este procedimiento (paso 5) dos veces más. Si los discos circulares de tarjeta FTA (micro-punch) se encuentran blancos (no se observa mancha rojiza de sangre) entonces puede omitirse el tercer lavado con buffer TE 10:1 y puede considerarse suficiente con dos lavados.

5.4 Amplificación por PCR multiplex

La amplificación de los 30 marcadores (INDELs) se realizó mediante la técnica de Reacción en Cadena de la Polimerasa (PCR) (Saiki et al., 1988). Cada volumen de reacción fue trabajado conforme a las especificaciones técnicas del manual Investigator-DIPplex-Handbook como se especifica a continuación.

Algunas consideraciones previas que se tomaron en cuenta:

- Se empleó puntas (tips) estériles con filtros hidrofóbicos.
- Se hizo vórtex a todos los tubos que contenían los componentes de la PCR.
- Breve centrifugado (5 segundos) después del vórtex a cada tubo componente de la PCR.

5.4.1 Condiciones para la PCR (Reacción en Cadena de la Polimerasa)

Tabla 4. Condiciones de PCR .* Contiene dNTPs , MgCl₂ y suero de albúmina bovino (BSA).

Componente	Volumen por reacción
Mix de racción A*	5.0 uL
Primer Mix	5.0 uL
Multi Taq2 DNA Polimerasa	0.6 uL
Agua libre de nucleasas	Variable
ADN	Variable
Volumen total	25 uL

5.4.2 Configuración de Termociclado

Se empleó un termociclador de Applied Biosystems® modelo Veriti de 96 posillos, la configuración del termociclado se muestra en la Tabla 5.

Tabla 5. Configuración de termociclado.* Temperatura de activación de la DNA polimerasa (Hot-Start).

Temperatura	Tiempo	Número de Ciclos
94°C*	4 min	1
94°C	30 s	30
61°C	120 s	
72°C	75 s	
68°C	60 min	1
10°C	∞	1

Se empleó como control positivo ADN de la línea celular 9948 y como control negativo agua libre de nucleasas, ambos vienen incluidos en el kit Investigator-DIPplex.

5.5 Electroforesis Capilar (EC)

Los fragmentos amplificados de ADN fueron genotipificados por electroforesis capilar (EC). Según los fundamentos explicados por Castagnino (2000) cabe precisar que la electroforesis capilar está basada en el movimiento diferencial de moléculas cargadas (iones) o no, por atracción o repulsión de un campo eléctrico generado por una fuente de alta tensión de 20 a 30 kv (Fig.3). Así, los fragmentos de ADN migran a través de un capilar de sílica cuyo diámetro interno va de 25 a 75 µm, la migración de moléculas a través de dicho capilar es monitoreado por un detector de absorbancia de alta sensibilidad en el cual se emplea longitudes de onda seleccionadas. Asimismo, la fuente de luz que se emplea para la detección

del flujo electroforético de estas moléculas proviene de un fotodiodo de referencia y un microenfocado óptico en la celda de detección o estuche óptico del capilar, consecuentemente la dispersión de la luz al contacto con las moléculas durante el flujo electroforético produce un patrón de picos gráficos en función de su movilidad electroforética, carga y masa, dicho patrón de picos gráficos es expresado en forma de electroferogramas (Fig.2).

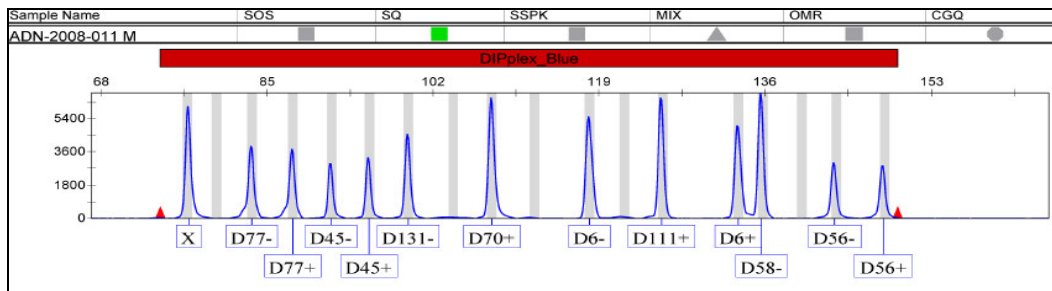


Figura 2. Electroferograma o perfil genético obtenido en equipo ABI 3500 de ocho capilares.

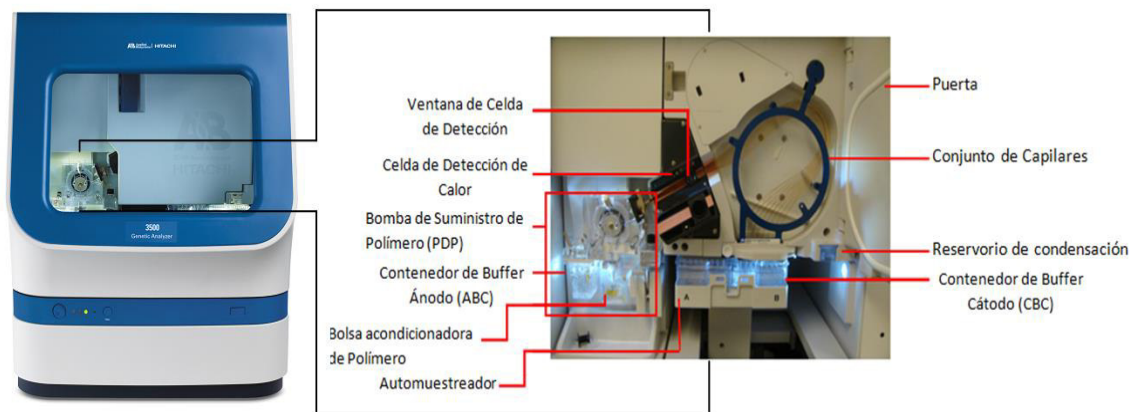


Figura 3: Analizador Genético modelo ABI 3500 de 8 capilares (Gráficos tomados del manual de Applied Biosystems® 3500/3500xL Genetic Analyzer Catalog Number 4405186, Publication Number 100031809, Revision A)

5.5.1 Materiales empleados en el Analizador Genético

Previo a las corridas electroforéticas de los fragmentos amplificados se realizó la calibración espectral del Analizador Genético conforme a las especificaciones técnicas del manual (Investigator DIPplex Handbook) para los 5 marcadores fluorescentes 6-FAM, BTG, BTY, BTR y BTO estos 5 marcadores fluorescentes son conocidos como BT5 y cada uno de ellos está asignado en el Analizador Genético con un color distintivo como se muestran en el Cuadro 2. El

procedimiento de la calibración espectral crea una matriz que es utilizado para una correcta superposición del espectro de fluorescencia para cada marcador.

Tabla 6. Consumibles propios del Equipo de Electroforesis Capilar ABI 3500

Material	Especificación
Capilar	8 Capilares de 36 cm de longitud de Analizador Genético ABI 3500 de Applied Biosystems.
Polímero	Polímero POP-4 para analizador genético ABI 3500 de Applied Biosystems.
Buffer	Contenedor de Buffer Ánodo (ABC) serie 3500 Contenedor de Buffer Cátodo (CBC) serie 3500

Cuadro 2. Marcadores fluorescentes de BT5.

COLOR	MATRIZ ESTANDAR
Azul (B)	6-FAM
Verde (G)	BTG
Amarillo (Y)	BTY
Rojo (R)	BTR
Anaranjado (O)	BTO

Los primers de los 30 marcadores (DIPlex Kit) empleados en este estudio están marcados por grupos de colores que se corresponden con 4 de estos 5 fluoróforos (Azul-6-FAM, Verde-BTG, Amarillo-BTY, Rojo-BTR). Asimismo, los 30 marcadores autosómicos fueron recodificados para su practicidad en los análisis bioinformáticos (Cuadro 3).

5.5.2 Preparación de la muestra para electroforesis capilar

Para realizar la electroforesis capilar, los fragmentos de ADN amplificados previamente se resuspendieron en una solución altamente desionizada denominada Hi-Di Formamida y DNA Size Standard BTO 550. Hi-Di Formamida actúa como un co-disolvente del ADN permitiendo una correcta y eficiente migración de los fragmentos amplificados en sistemas de electroforesis capilar como el Analizador Genético ABI 3500 de 8 capilares de Applied Biosystems, mientras el DNA Size Standard BTO 550 constituye una regla o carril interno destinado a ser utilizado en la asignación de valores a los fragmentos (picos de electroferograma) detectados por fluorescencia en el sistema de electroforesis capilar, DNA Size standard BTO 550 se usa comúnmente en genotipificación y análisis de fragmentos de ADN. Se emplearon volúmenes de Hi-Di Formamida y DNA Size Standard conforme a las especificaciones técnicas del manual investigator DIPplex handbook, (Tabla 7).

Cuadro 3. Matriz Estándar y correlación con cada locus y códigos asignados

N°	MATRIZ ESTANDAR	COLOR	DIP-LOCUS
Det. Sexo	6-FAM	AZUL	Amelogenin X
Det. Sexo			Amelogenin Y
1			HLD77
2			HLD45
3			HLD131
4			HLD70
5			HLD6
6			HLD111
7			HLD58
8			HLD56
9	BTG	VERDE	HLD118
10			HLD92
11			HLD93
12			HLD99
13			HLD88
14			HLD101
15			HLD67
16	BTY	AMARILLO	HLD83
17			HLD114
18			HLD48
19			HLD124
20			HLD122
21			HLD125
22			HLD64
23			HLD81
24	BTR	ROJO	HLD136
25			HLD133
26			HLD97
27			HLD40
28			HLD128
29			HLD39
30			HLD84

Tabla 7. Componentes y volúmenes de reactivos para electroforesis capilar.

COMPONENTE	VOLUMEN POR MUESTRA
Hi-Di-Formamida	12.0 μ L
DNA-Size Standard (BTO)	0.5 μ L

Los pasos seguidos previos a la inyección electroforética fueron los siguientes:

1. Para cada muestra a analizar alicuotar 12 μ L de la mezcla (Hi-Di-Formamida y DNA-Size Standard) en cada posillo.
2. Añadir 1 μ L del producto de PCR. Cabe precisar que el manual Investigator DIPplex recomienda inyectar el ladder alélico cada 3 inyecciones de muestra, es decir, cada 24 muestras (si está empleando un Analizador Genético de 8 capilares) para una mejor asignación de valores de los fragmentos a analizar.
3. Centrifugar brevemente la placa para llevar todo el líquido inyectado al fondo de cada posillo, evitar presencia de burbuja o que el líquido se encuentre al costado o a la mitad de cada posillo (Fig.4).

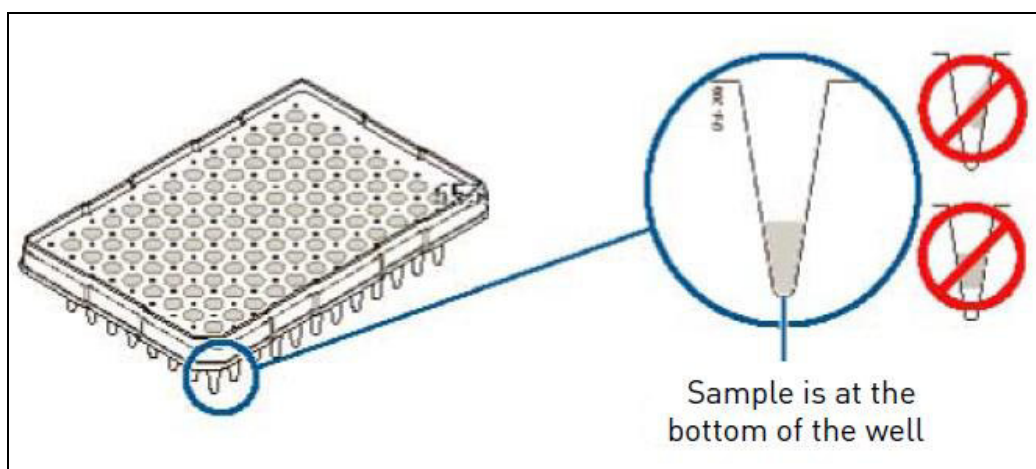


Figura 4. Muestra posición adecuada de componentes líquidos inyectados. Investigator DIPplex Handbook 2013.

4. Denaturar la placa por 3 minutos a 95 °C.
5. Enfriar la placa en hielo o refrigeración durante 3 minutos.
6. Cargar la placa al Analizador Genético.

Los parámetros configurados y aplicados en el Analizador Genético ABI 3500 de Applied Biosystems se muestran en el Cuadro 4, conforme a las especificaciones técnicas del manual de Investigator DIPplex.

Cuadro 4. Parámetros configurados para ejecutar electroforesis capilar de 30 marcadores (DIPlex Kit) en Analizador Genético ABI 3500 de Applied Biosystems

PARÁMETRO	CONFIGURACIÓN
Tipo de aplicación	HID
Longitud de capilar	36 cm
Polímero	POP-4
Dye Set	BT5
Módulo de corrida	HID36_POP4
Nombre de Protocolo	CORRIDA 230914 DIPLEX PLACA#
Temperatura (°C)	Default
Voltage de Corrida (kV)	Default
Voltage de pre-corrida (kV)	Default
Voltage de Inyección (kV)	3.0
Tiempo de Corrida (s)	900
Tiempo de pre-corrida (s)	Default
Tiempo de Inyección (s)	8.0
Data Delay (s)	Default
Opciones Avanzadas	Default

Una vez creado el protocolo de corrida electroforética con los parámetros que se muestran en la Tabla 11, se confirma y guarda los cambios en el Software 3500 Series Data Collection del Analizador Genético.

También es necesario configurar los parámetros en el mismo software para el Size Standard dirigiéndose a la opción “Librería” seleccionar “Análisis” seguido de “Size Standard” dar clic en “Crear” (se precisa que todas las opciones estan en Inglés). En ese sentido el Size Standard 550 (BTO) deberá configurarse y ser usado con los siguientes fragmentos de longitud: 60, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 250, 260, 280, 300, 320, 340, 360, 380, 400, 425, 450, 475, 500, 525, y 550 pares de bases (Fig.5).

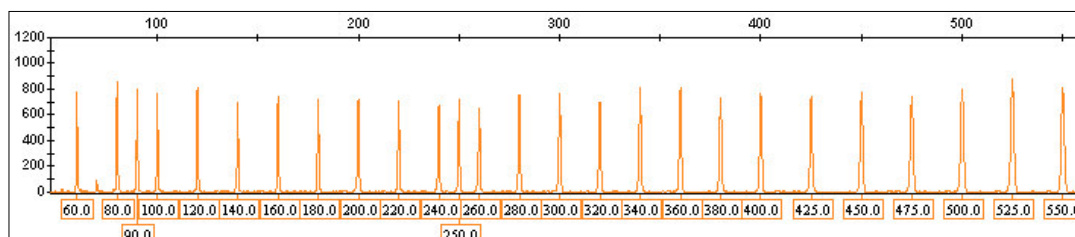


Figura 5. Electroferograma del DNA Size Standard 550 (BTO), con sus respectivos fragmentos de longitud en pares de bases.

5.5.3. Obtención de datos de electroforesis capilar

La obtención de datos de los fragmentos analizados mediante electroforesis capilar se realizó empleando el software GeneMapper ID-X, propio del Analizador Genético ABI3500 de Applied Biosystems, el cual está diseñado específicamente para aplicaciones de identificación humana, adicionalmente para estudios de genética de poblaciones mediante análisis de fragmentos de ADN y/o secuenciamiento de ADN y permite la asignación de alelos con alta precisión.

Por otro lado, los archivos molde o plantillas que provee el kit de Investigator DIPlex para el análisis de los 30 marcadores de inserción y delección son archivos complementarios y compatibles con el software GeneMapper ID-X que permite simplificar el análisis de datos, al utilizarse ambos conjuntamente, permiten asignar los alelos relativo a la longitud de los STRs. Opcionalmente tambien corresponden los picos del electroferograma con los fragmentos de longitud en pares de bases o también con las unidades relativas de fluorescencia (RFU). Estos datos (genotipificados) son tabulados y exportados para luego ser analizados por diferentes programas bioinformáticos.

Los procedimientos técnicos de configuración de software, definición de Size Standard, ajustes de archivo molde Investigator DIPlex, ajustes de paneles de bins propios del análisis y obtención de datos en el Analizador Genético ABI 3500 de Applied Biosystems, se realizaron conforme a las especificaciones de la guía “Investigator Template Files for Gene Mapper ID-X Software User Guide” que se puede descargar libremente de la página web “www.qiagen.com”, así como otras herramientas que permiten un análisis práctico y relativamente sencillo de los 30 marcadores de INDELs.

5.6 Análisis de datos

Antes de exportar los datos a formatos que permitan su análisis genético, se ha revisado la proporción de altura de los picos (RFU) en todos los genotipos, la presencia de alelos nulos o posibles mutaciones, la ubicación correcta de los picos en los bins (regla alélica o “allellic ladder), es decir, la calidad de los electroferogramas. Esta etapa es muy importante e imprescindible ya que de ella depende los resultados y las inferencias que se realizarán como parte del estudio.

Los datos obtenidos de los electroferogramas fueron analizados mediante el programa Gene Mapper ID-X, (fig. 6), exportados como tablas en formato “.txt” delimitado por tabulación. Una vez guardado los archivos en formato “.txt” fueron convertidos a archivos Excel para recodificar cada uno de los 30 marcadores DIPlex (Fig.7) con la finalidad de facilitar su análisis en otros programas bioinformáticos.

Sobre los valores asignados a cada marcador, las inserciones fueron asignados con número 2 (dos) y las deleciones con el número 1 (uno), así por ejemplo; si un marcador presenta ambos (inserción y delección) 2/1 respectivamente, en homocigosis 2/2 o 1/1.

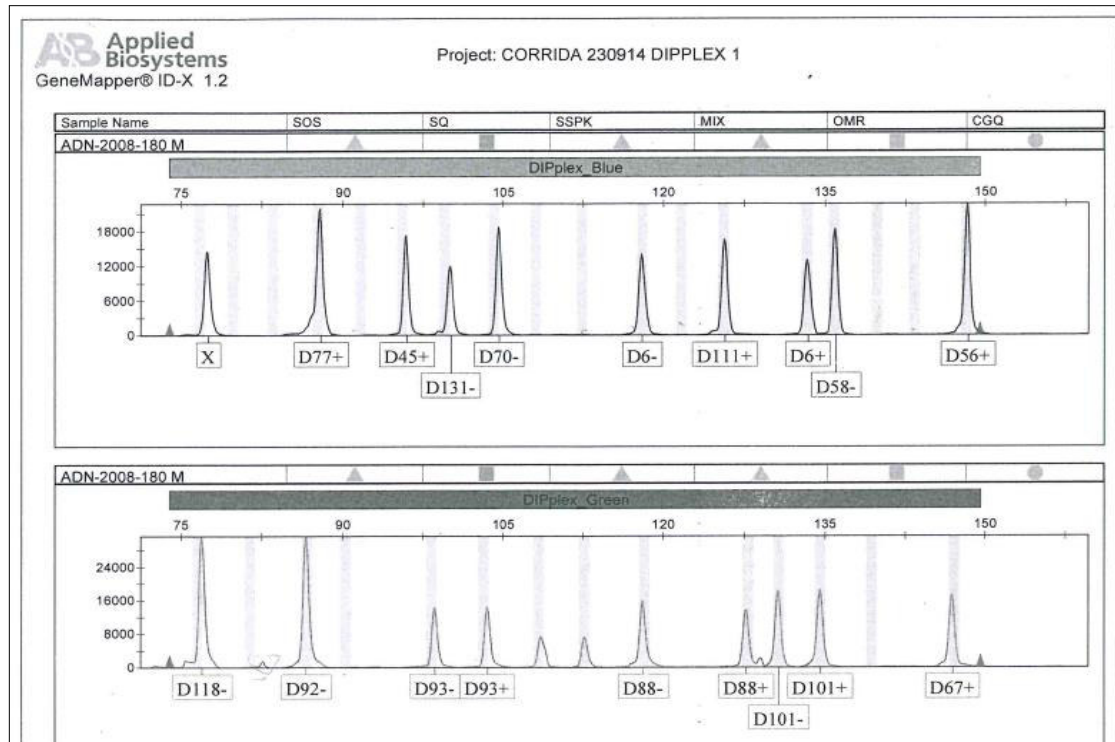


Figura 6. Electroferograma de una muestra obtenido en analizador genético ABI 3500 de Applied Biosystems mediante programa Gene Mapper ID-X

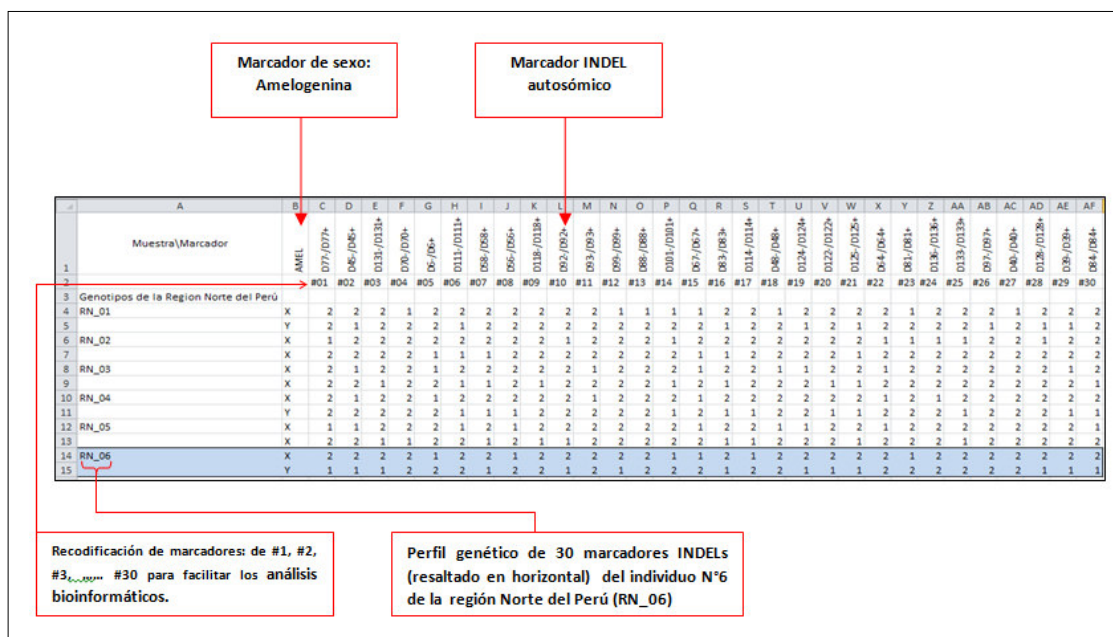


Figura 7. Muestra los datos recodificados de perfiles genéticos INDELs en formato Excel.

5.7. Análisis Genético Poblacional

5.7.1 Análisis de Estructura Poblacional

El análisis de estructura poblacional se realizó utilizando el programa Structure V. 2.3.4 (Pritchard., 2009). Para ello, a fin de observar la existencia o no de algún patrón de agrupamiento con relación a la distribución geográfica de los genotipos, fueron agrupados según su localización en Norte, Centro y Sur con el siguiente detalle: Norte con 55 individuos procedentes de Piura, Tumbes, Amazonas, San Martín, Lambayeque, La Libertad y Cajamarca. Centro con 59 individuos procedentes de Ancash, Norte de Lima, Lima, Ica, Huánuco, Ucayali, Junín y Cerro de Pasco. Sur con 22 individuos procedentes de Ayacucho, Cuzco, Puno y Arequipa (Fig.8).

Los parámetros de configuración fueron los siguientes: 10000 generaciones de MCMC (Método de Cadenas de Markov Montecarlo) y Burn-in de 5000. El modelo de ancestría analizado fue Admixture. El modelo de frecuencia alélica: Correlativo, los valores de K simulados de 1 a 10 ($K = 1 - 10$) con 8 iteraciones en cada análisis para mayor consistencia en los resultados.

Los datos ingresados en el programa fueron los siguientes: 136 genotipos INDELs, dos filas por genotipo (alelo 1 y alelo 2), 30 loci con número de ploidias igual a 2, y los espacios en blanco (Missing Date) representados por -9, asimismo en el genotipado el valor 1 corresponde a delección y el valor 2 corresponde a inserción y en el archivo de entrada las primeras 5 columnas

corresponden a: primero identificador de muestra (Label), segundo identificador de población (Pop-ID), información de población (Flag), cuarto y quinto representan columnas extra (Columnas extra igual a 2).

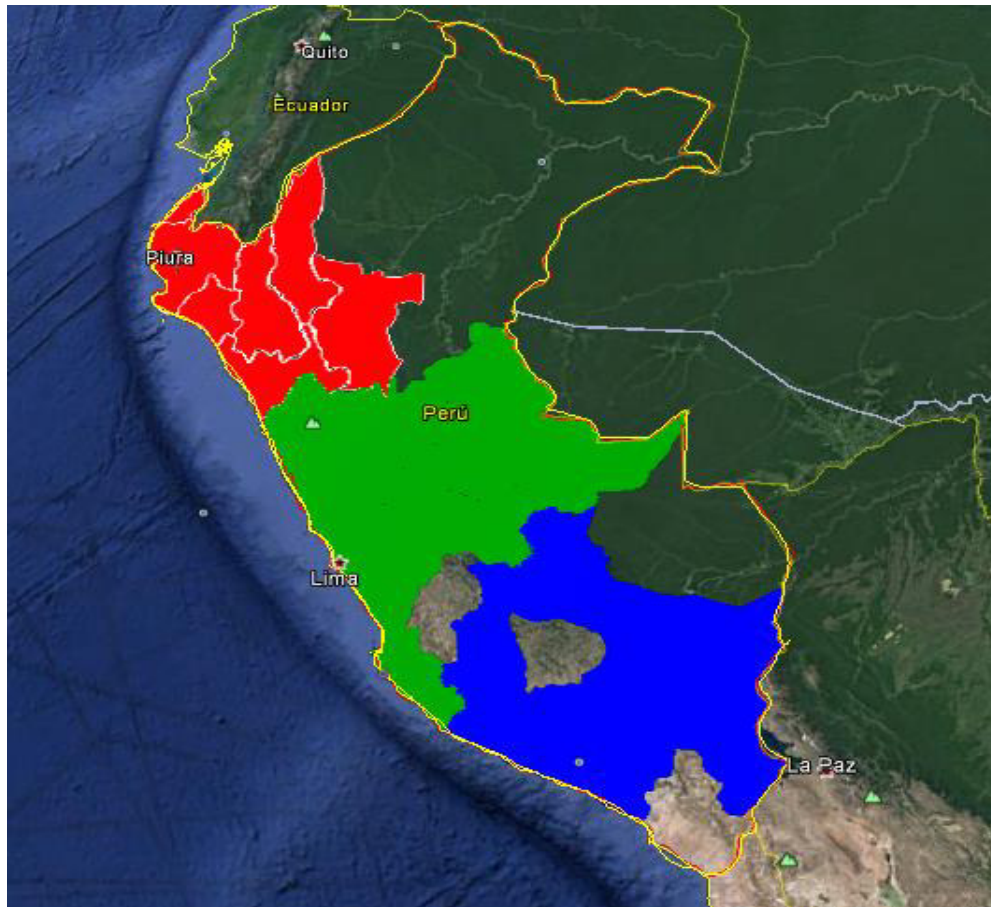


Figura 8. Mapa del Perú con las tres regiones estudiadas. En rojo la región Norte, en verde la región Centro, en azul la región Sur. (Maps GoogleEarthPro.<https://earth.google.com>)

5.7.2 Análisis de Estructura Poblacional comparado con datos de poblaciones europeas y sudamericanas

A fin de observar la estructura genética a una escala intercontinental, en este caso entre Europa (España Central y País Vasco) y América del Sur (Perú y Uruguay), con el mismo set de marcadores analizamos 398 perfiles genéticos INDELs agrupados de la siguiente manera: 136 de Perú propios del presente estudio, 131 de Uruguay (Saiz et al., 2014), 131 Europeo [71 España-Central y 60 País Vasco] (Martin et al., 2013), los perfiles genéticos de las poblaciones de Uruguay, España Central y País Vasco fueron amablemente compartidos por los autores de cada investigación los cuales fueron solicitados con la finalidad de observar la estructura poblacional entre dichos países y la población peruana, principalmente

por los antecedentes históricos de colonización por parte de España, tanto en Perú como en Uruguay, durante el Siglo XVI y Siglo XVII.

La configuración del programa se realizó con los siguientes parámetros: 100000 generaciones de MCMC y 50000 de burn-in, Modelo de ancestría: Admixture, Modelo de Frecuencia Alélica: Correlativo, ($K = 1-4$) y se realizó 10 iteraciones en cada caso para darle consistencia a los resultados.



Figura 9. Muestra el mapa de América del Sur (Perú y Uruguay) y una parte del continente europeo (España Central y País Vasco), con los cuales se hizo el análisis de estructura poblacional a escala intercontinental, (Maps Google Earth Pro. EarthPro.<https://earth.google.com>)

5.7.3 Prueba de Equilibrio de Hardy–Weinberg

El modelo genético poblacional del equilibrio de Hardy–Weinberg propuesto en 1908 por G.H. Hardy y W. Weinberg (Thompson & Guo, 1992) quienes demuestran por separado que “en una población panmictica (es decir, con apareamientos al azar), de gran tamaño y donde los individuos son igualmente viables y fecundos, el proceso de herencia genética, por sí mismo no cambia las frecuencias alélicas ni genotípicas de un determinado locus” (Cabrero & Camacho, 2011). En el presente estudio asumiremos (hipótesis nula) que las poblaciones en estudio se encuentran en equilibrio de Hardy–Weinberg (basándonos en las características de este modelo), sin embargo, en caso de constatar desviaciones significativas al Equilibrio de Hardy–Weinberg podremos concluir que no se encuentra en equilibrio de Hardy–Weinberg, que no se ajusta a dicho modelo y que los alelos estudiados poseen un patrón distinto de herencia.

Para realizar la prueba de Equilibrio de Hardy-Weinberg empleamos el programa Arlequin. V.3.5. (Excoffier, 2010). Para el test del Equilibrio de Hardy-Weinberg se configuró con 1000000 de valor de cadenas de Markov y 100000 pasos de dememorización, el análisis fue realizado locus por locus.

5.7.4 Prueba Fst

La prueba de distancia genética de a pares entre poblaciones, en inglés “Population Pairwise Genetic Distances (Fst)” calcula el Fst para todos los pares de poblaciones, así como los diferentes índices de disimilaridad (distancia genética) entre los pares de poblaciones (Reynolds et al. 1983; Slatkin, 1995). La significancia de la distancia genética se prueba por permutación de haplotipos o individuos entre poblaciones (Excoffier, 2010). Para medir los valores de Fst se puede hacer mediante los siguientes modelos:

Distancia de Slatkin Cálculos de Slatkin

Considera un simple modelo demográfico donde dos poblaciones haploides de tamaño “N” divergieron hace “t” generaciones a partir de una población de tamaño idéntico. Estas dos poblaciones han permanecido aislados desde entonces sin intercambiar migrantes. Bajo estas condiciones, el parámetro Fst puede ser expresado en términos de tiempos de coalescencia “t₁” que es el tiempo de coalescencia media de dos genes extraídos de dos poblaciones diferentes, y el “T₀” que es el tiempo de coalescencia media de dos genes extraídos de la misma población. Utilizando el análisis de varianza el Fst puede expresarse como:

$$F_{ST} = \frac{\bar{t}_1 - \bar{t}_0}{\bar{t}_1} \quad (\text{Slatkin, 1991})$$

Si el t₀ es igual al número de generaciones “N” y el t₁ es igual a t + N generaciones, con estas premisas la ecuación se expresaría de la siguiente manera:

$$F_{ST} = \frac{\tau}{\tau + N} \quad \text{Por lo tanto la distancia se expresaría así:}$$

$$D = F_{ST} / (1 - F_{ST})$$

la cual es igual a t / N, de esta ecuación se deduce que si aplicamos el modelo de Fst de Slatkin la distancia genética es proporcional al tiempo de divergencia entre dos poblaciones.

- **Distancia de Reynolds**

Reynolds y colaboradores en 1983 linearizan los valores F_{ST} para tiempos de divergencia cortos. En este caso el parámetro F_{ST} entre pares de poblaciones haploides estacionarias de tamaño N que hayan divergido hace “ t ” generaciones se expresa en la siguiente ecuación:

$$F_{ST} = 1 - \left(1 - \frac{1}{N}\right)^t \approx 1 - e^{-t/N} \quad (\text{Reynolds et al., 1983})$$

La distancia genética “ D ” se expresa como: $D = -\log(1 - F_{ST})$

Esta ecuación resulta aproximadamente proporcional a t/N para tiempos de divergencia cortos.

- **Diferencia de a pares**

Calcula el número promedio de diferencias de a pares de individuos o haplotipos dentro y entre poblaciones (Nei & Li., 1979). Como una distancia genética adicional entre poblaciones adiciona un parámetro “ D ” que representa el número de diferencias nucleotídicas entre poblaciones, así tenemos D y D_A que se calculan para la población 1 y 2, respectivamente, como se muestra en las siguientes ecuaciones:

$$D = \hat{\pi}_{12} = \sum_{i=1}^k \sum_{j=1}^{k'} x_{1i} x_{2j} \delta_{ij} \quad , \quad y \quad D_A = \hat{\pi}_{12} - \frac{\hat{\pi}_1 + \hat{\pi}_2}{2}$$

Donde K y K' es el número de haplotipos distintos en la población 1 y 2 respectivamente, x_{1i} es la frecuencia del i -th haplotipo en la población 1, y δ_{ij} es el número de diferencias entre el haplotipo i y el haplotipo j . Bajo esta misma notación, con respecto al tiempo de coalescencia, el número de diferencias nucleotídicas esperadas para D_A es:

$$D_A = 2\mu (\bar{t}_1 - \bar{t}_0) = 2\mu\tau$$

Donde μ es el promedio de la tasa de mutación por nucleótido, τ es el tiempo de divergencia entre dos poblaciones. Así se espera que D_A tienda a incrementar linealmente con el tiempo de divergencia entre poblaciones.

Los valores de F_{ST} , así como otros parámetros asociados a este cálculo, como el número de permutaciones, grado de significancia, distancia euclidiana, matriz de distancias, y los F-estadísticos fueron calculados con el programa Arlequin. V.3.5. (Excoffier, 2010).

5.8. Índices de diversidad estándar

5.8.1. Diversidad Genética. La diversidad Genética es equivalente a la Heterosigosis esperada para datos diploides. Está definido como la probabilidad de que dos haplotipos escogidos aleatoriamente sean diferentes en una muestra poblacional. La diversidad genética y su varianza de muestreo se estimaron mediante las siguientes ecuaciones:

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$
$$V(\hat{H}) = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\sum_{i=1}^k p_i^3 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right\}$$

Donde “**n**” es el número de copias de genes en la muestra, “**k**” es el número de haplotipos y “**p_i**” es la frecuencia de la muestra en el “**i-th**” haplotipo.

Estos cálculos se realizaron con el programa Arlequin. V.3.5. (Excoffier, 2010) en el cual están implementados estos algoritmos a fin poder determinar diferentes índices de diversidad genética que a continuación se explican:

5.8.2. Cálculo de Heterocigosidad Esperada y Heterocigosidad Observada

La heterocigosidad esperada (He) o índice de diversidad de Nei (1973), se estimó con el programa Arlequin V.3.5 (Excoffier, 2010), calculado como la fracción de individuos que podrían ser heterocigotos para cualquier locus tomado al Azar. La He difiere de la Heterocigosidad Observada (Ho) en que la primera se calcula como una predicción basada en el conocimiento de la frecuencia alélica del total de muestras de individuos estudiados. Por lo tanto explicaremos en los resultados si hubieran desviaciones de la Ho respecto de la He, lo cual constituiría un indicador de la dinámica poblacional según el principio de Equilibrio de Hardy-Weinberg (1908).

El programa Arlequin V.3.5. tiene implementado la ecuación de Nei (1987) que permite calcular la heterocigosidad esperada, dicha ecuación se expresa de la siguiente manera:

$$\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right) \quad (\text{Nei, 1987})$$

Donde “n” es el número de copias de genes, “k” es el número de haplotipos, (p_i) es la frecuencia de la muestra para el haplotipo “i-th”.

5.8.3. Estimación de Frecuencias Alélicas. Las frecuencias alélicas se calcularon empleando el programa Arlequin V.3.5 (Excoffier, 2010). Estos cálculos se realizan según el número de individuos que portan un determinado número de alelos (por ejemplo los alelos A1 y A2), como se muestra a continuación:

$$p = \frac{P + \frac{1}{2}H}{N} : \quad X \quad \text{Siendo } X + Y = 1,$$

$$q = \frac{Q + \frac{1}{2}H}{N} : \quad Y$$

Donde: “p” es la frecuencia alélica para el alelo A1 , “q” es la frecuencia alélica para el alelo A2, “P” es la frecuencia genotípica de A1 A1, “H” es la frecuencia genotípica “A1A2” mientras “Q” es la frecuencia genotípica “A2 A2”.

5.8.4. Análisis de Varianza Molecular (AMOVA)

Realizamos estos cálculos para evaluar un marco de referencia de la estructura genética a partir de la información contenida en las frecuencias alélicas, la cual es introducida en una matriz de distancias cuadradas euclidianas y a partir de ella se obtienen los componentes de covarianza asociado con posibles niveles de estructura genética (Casalla, 2003)

Asimismo para estimar los índices de estructura genética o agrupamiento usamos la información de contenido alélico de los genotipos y sus frecuencias (Excoffier et al. 1992). Para probar la proporción de la variación genética atribuible a la variación dentro de las poblaciones, entre las poblaciones, y entre los grupos poblacionales en una región (González., 2011). Este análisis también puede utilizarse para determinar la distribución relativa de la diversidad genética entre grupos, entre poblaciones dentro de grupos o dentro

de las poblaciones, en ese sentido la estructura jerárquica es frecuentemente cuantificada a través del AMOVA (Allendorf & Luikart., 2007)

Se hizo el análisis de Varianza Molecular (AMOVA) basado en permutaciones de las distancias genéticas comparando tres regiones poblacionales (Norte, Centro y Sur) del Perú, evaluando la hipótesis nula de uniformidad genética entre pares de poblaciones para ello utilizamos el programa Arlequin V.3.5 (Excoffier, 2010).

5.8.5. Estimación de Desequilibrio de Ligamiento

Se realizó estos cálculos para evaluar la presencia de asociación significativa entre pares de loci basado en el test de desequilibrio de ligamiento. Este test se puede realizar con todos los tipos de datos genéticos excepto con los datos de frecuencias, si hubiera menos de dos loci polimórficos no tendría sentido realizar esta prueba. El procedimiento es análogo al test de Fisher que evalúa la probabilidad de conteo de genotipos que es igual o menos probable al conteo de genotipos observados (Slatkin, 2008).

5.9. PARÁMETROS ESTADÍSTICOS DE INTERÉS FORENSE

Es importante estimar los parámetros estadísticos de interés forense, ya que nos indican la utilidad a priori esperada para cada locus y para el sistema genético en conjunto. Estos son:

5.9.1. Probabilidad de Match (MP)

La probabilidad de coincidencia también es conocida como probabilidad de match (Jones ,1972) y se define como la probabilidad de que dos individuos elegidos al azar tengan idénticos genotipos en uno o más sistemas genéticos, es decir, es la probabilidad de que dos individuos no relacionados compartan un mismo perfil de ADN. Esta probabilidad puede aumentar cuando tenemos un perfil de ADN parcial por degradación o cuando tenemos un reducido número de locus informativos. Además, también puede aumentar si el sospechoso y el culpable real son emparentados, o si se originan de la misma subpoblación. El único caso excluible a estos valores es el de los gemelos monocigóticos, ya que comparten el mismo material genético. Los cálculos de este parámetro se realizaron en el programa Power Stats V12.Xls® (Promega Corporation, Madison, WI. USA).

5.9.2. Poder de Discriminación (PD)

El poder de discriminación (PD) se caracteriza por la probabilidad de que dos individuos no relacionados y seleccionados al azar, puedan diferenciarse genéticamente mediante el análisis de un marcador o un conjunto de marcadores (Smalldon & Moffat, 1973). El PD depende del número de locus analizados y del polimorfismo de cada uno y de sus frecuencias alélicas. El poder de discriminación para este estudio fue calculado utilizando el programa *Power Stats V12.Xls®* (Promega Corporation, Madison, WI. USA).

$$PD = 1 - \sum P_{ij}^2$$

Donde P_{ij} = Frecuencia Genotípica.

5.9.3. Contenido de Información Polimórfica (PIC)

El valor del contenido de información polimórfica fue descrito por Botstein (1980) es un índice que evalúa la normatividad de un marcador en la población de acuerdo con las frecuencias de los alelos. Según la clasificación de Botstein los marcadores con valores de PIC superiores a 0,5 son considerados muy informativos, con valores entre 0,25 y 0,50 medianamente informativos y con valores inferiores a 0,25, poco informativos. Un microsatélite es considerado como un buen marcador molecular cuando presenta un número de alelos superior a 5 y el PIC con índices por encima de 0,75 (Foissac *et al.*, 1997). El contenido de información polimórfica se obtuvo utilizando el programa *Power Stats V12.Xls®* (Promega Corporation, Madison, WI. USA).

5.9.4 Poder de Exclusión (PE)

El poder de exclusión se define como la probabilidad de excluir un individuo que esté involucrado en algún caso forense. En casos de paternidad, por ejemplo, sería la probabilidad de excluir un “falso padre” (Jamieson, 1994; Lee *et al.*, 1998; Krenke *et al.*, 2002). Este parámetro nos permite establecer la proporción de individuos falsamente implicados en un peritaje.

El Grupo Español y Portugués de la Sociedad Internacional de Genética Forense (GEP-ISFG) considera que para que un Laboratorio sea competente en la realización de estas pruebas, estos índices deben estar por encima del 99,9% (Rosado., 2008). El poder de Exclusión fue calculado utilizando el programa *Power Stats V12.Xls®* (Promega Corporation, Madison, WI. USA).

5.9.5. Índice Típico de Paternidad (TPI)

Este índice evalúa la información de un marcador en la población de acuerdo a las frecuencias de sus alelos, su medición es similar al valor de la heterosigosidad y oscila entre 0 y 1. Se obtiene de multiplicar la probabilidad de cada posible cruzamiento (estimado a partir de las frecuencias alélicas) por la probabilidad de que sean informativos, es decir, que se pueda identificar al progenitor del que procede el alelo. El Índice Típico de paternidad (TPI) fue calculado utilizando el programa Power Stats V12.Xls® (Promega Corporation, Madison, WI. USA).

$$PIC = 1 - \sum_{i=1}^n (p_i^2) - 2 \sum_{i=1}^{n-1} \sum_{j=1-i}^n p_i^2 p_j^2 \quad (Brenner \& Morris, 1990)$$

Donde p_i hasta p_n , son las frecuencias de los n alelos.

6. ANÁLISIS DE ESTRUCTURA POBLACIONAL

Este es un análisis basado en modelos de agrupamiento para inferir la estructura poblacional utilizando datos genotípicos de marcadores no ligados. El método fue implementado en un artículo por Pritchard y colaboradores (2009).

La aplicación de este método permite demostrar la presencia o no de estructura poblacional, identificando distintas poblaciones genéticas, asignando individuos a las poblaciones según su perfil genético, también permite identificar migrantes e individuos mestizos, siempre y cuando se esté trabajando con poblaciones de referencia (etnias o sus equivalentes). En los análisis se asume hipotéticamente un modelo en el cual existen “K” poblaciones (donde el valor de K puede ser desconocido y se desea conocer en la población analizada), cada uno de los cuales es caracterizado por un set de frecuencias alélicas por cada locus. Los individuos en las muestras son asignados probabilísticamente a poblaciones o junto a dos o más poblaciones en los casos cuyos genotipos indican que se encuentran mezclados. Asimismo, se asume que dentro de las poblaciones analizadas todos los loci se encuentran en Equilibrio de Hardy-Weinberg y en desequilibrio de ligamiento, es decir, los alelos que se transmiten de progenitores a descendientes son copias exactas de los progenitores o en su defecto presentan una baja tasa de recombinación (por debajo del 50%).

Los análisis de estructuración poblacional no asumen un proceso particular de mutación y se puede aplicar a una variedad de marcadores moleculares como son: Los microsatélites (STRs), SNPs, RFLPs e INDELs. El modelo asume que todos los marcadores mencionados no se encuentran en desequilibrio de ligamiento dentro de subpoblaciones, por lo tanto no

podemos manipular marcadores que estén extremadamente juntos o cercanos en un cromosoma (Pritchard et al., 2009).

Para los análisis de estructura poblacional empleamos en programa Structure Structure v 2.3.4 (Pritchard et al., 2009). Se empleó 30 marcadores bialélicos de uso forense de Inserción y Delección INDELs (Investigator DIPLEX-Kit, Germany) en 136 peruanos mestizos.

7. RESULTADOS

7.1. Análisis de Estructura Poblacional en tres regiones del Perú

Los análisis de estructura poblacional muestran que el agrupamiento de las poblaciones para valores de K analizados de 1 a 10 muestran un mejor agrupamiento para K igual 3. Sin embargo se observa que los patrones de distribución de genotipos en cada grupo son muy similares, es decir, no muestra diferencias significativas entre Norte, Centro y Sur del Perú en cuanto a estructuración poblacional (Fig. 10). Dado que en el presente estudio no empleamos datos de poblaciones de referencia para ancestría (genotipos de etnias u otras poblaciones que habrían contribuido al acervo génico de la población analizada), únicamente buscamos observar la existencia o no de evidencias de estructuración o agrupamiento en base a los genotipos analizados.

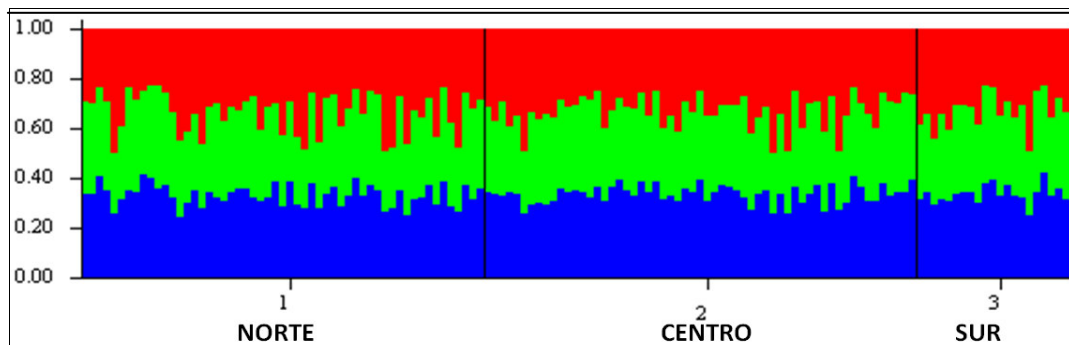


Figura 10. Resultados del análisis de estructuración poblacional de las tres regiones del Perú

Se puede observar (Fig.10) la distribución de genotipos en el Norte, Centro y Sur del Perú con un patrón similar en las tres regiones (rojo, verde y azul). Nuestro análisis evidencia que no se observa niveles de estructuración o agrupamiento de la población mestiza en el Perú analizados en grupos de Norte, Centro y Sur. Esto podría deberse a la no existencia de diferencias significativas en la diversidad

genética entre las poblaciones analizadas, es decir, podría tratarse para los efectos de análisis genético poblacional a los 136 genotipos como una única población.

Por otro lado no se descarta la posibilidad de que nuestros resultados podrían tener un efecto de sesgo por tamaño muestral, los cuales, siendo 136 genotipos, podrían representar, como también no, genéticamente a la población peruana, ello podría corroborarse en otros estudios incrementando el tamaño muestral y observando si dichos resultados varían.

Tabla 8. Número de agrupamientos con mayor verosimilitud ensayado con 136 genotipos.

K	Reps	Media LnP(K)	Desv. Estand.		Ln''(K)	Delta K
			LnP(K)	Ln'(K)		
1	3	-4765.4667	0.1528	NA	NA	NA
2	3	-4904.8667	81.9235	-139.4	114.933333	1.402934
3	3	-4929.3333	30.5873	-24.466667	55.166667	1.80358
4	3	-5008.9667	166.7903	-79.633333	4.866667	0.029178
5	3	-5083.7333	288.3672	-74.766667	118.866667	0.412206

Se puede observar (Tabla 8) el mejor valor probabilístico de K, asignado mediante el método de Evanno (Evanno et al., 2005), muestra para valores ensayados de 1 al 5, que la verosimilitud promedio de las posibles “K” poblaciones resulta consistente en tres poblaciones (K = 3).

7.2. Análisis de Estructura Poblacional comparado con datos de poblaciones europeas y sudamericanas

Es de esperar que la variación de la diversidad genética dentro de una misma población como la peruana resulte más homogéneo que al compararlo con una población más distante geográficamente como la europea y eso es lo que se obtuvo al comparar con la población española (Fig.11). Si bien es cierto que dos poblaciones de América del Sur como la peruana y uruguaya están más cerca geográficamente que con la europea, nuestros resultados parecerían contradictorios ya que la población uruguaya no se agrupa con la población peruana sino más bien con la población española.

En ese sentido, los resultados del análisis de estructura poblacional de 398 genotipos procedentes de España Central (71 personas), País Vasco (60 personas), Uruguay (131 personas) y Perú (136 personas) muestran que los genotipos de Perú forman un grupo poblacional (en verde) que se distingue de la población de España Central, País Vasco y Uruguay (en rojo) (Fig. 11). Se observa en este

análisis que el mejor valor probabilístico de K, asignado mediante el método de Evanno (Evanno et al. 2005) es igual a 2 ($K=2$).

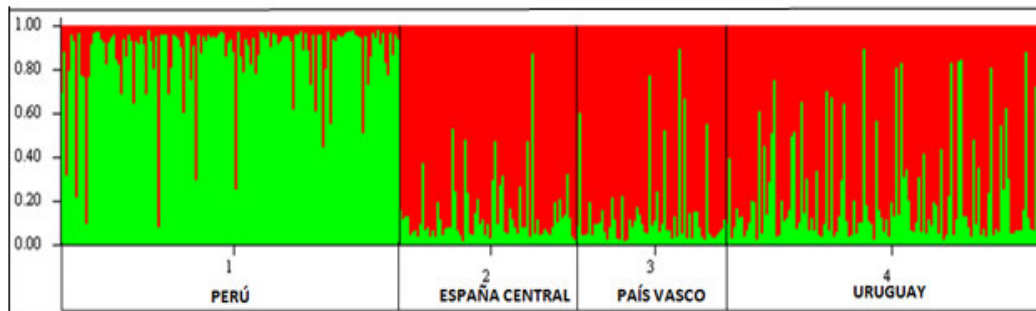


Figura 11. Resultados del Análisis de estructura poblacional a nivel intercontinental. Se muestra principalmente dos grupos, donde Uruguay muestra bastante similitud con la composición genética de España Central y País Vasco, mientras que Perú muestra una composición genética que se distingue de estas poblaciones.

Tabla 9. . Número de agrupamientos con mayor verosimilitud ensayado a nivel intercontinental

K	Reps	Media. LnP(K)	Desv. Estand. LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	10	-15925.030000	0.105935	NA	NA	NA
2	10	-15508.300000	4.088738	416.730000	559.870000	136.929784
3	10	-15651.440000	31.620148	-143.14000	64.230000	2.031300
4	10	-15730.350000	51.644710	-78.910000	69.990000	1.355221

Se observa (Tabla 9) el mejor valor de K encontrado ($K=2$) simulado para valores de K de 1 a 4 con 100000 Repeticiones de MCMC. Muestra dos agrupamientos: el primero conformado por genotipos peruanos y el segundo conformado por genotipos de España Central, País Vasco y Uruguay. Otro aspecto que podemos resaltar de este análisis es que la estructura genética de las poblaciones no siempre se ve reflejado en la proximidad geográfica de los individuos, así tenemos que siendo Uruguay y Perú países sudamericanos que históricamente fueron colonizados por España, sin embargo el Perú muestra una composición propia que se va distinguiendo de España Central y País Vasco, así como de Uruguay, es decir, si bien es cierto no logramos observar diferenciación o agrupamiento dentro la población peruana (Fig.10), pero, al realizar los análisis a nivel intercontinental (mayor distancia) ya podemos observar diferencias o un grado de agrupamiento ($K=2$).

7.3. Prueba de equilibrio de Hardy–Weinberg

No se observó desviaciones al equilibrio de Hardy-Weingberg, excepto en los marcadores HLD45, HLD131, HLD67 cuyos P-values fueron 0.04704, 0.03641 y 0.01789 respectivamente en la población peruana. Sin embargo, luego de realizar la corrección de Bonferroni ($0.05/30 = 0.001666$) se observó que ninguno de los 30 marcadores mostraron desviaciones al Equilibrio de Hardy-Weinberg.

Tabla 10. Resultados del análisis de Equilibrio de Hardy-Weinberg

N° Locus	Locus	(H.o)	(H.e)	P-value	Desv.Est.
1	HLD77	0.52874	0.46774	0.25239	0.00045
2	HLD45	0.33333	0.42608	0.04704	0.00022
3	HLD131	0.27586	0.36835	0.03641	0.00019
4	HLD70	0.45977	0.45445	1.00000	0.00000
5	HLD6	0.48276	0.47891	1.00000	0.00000
6	HLD111	0.43678	0.49485	0.28391	0.00045
7	HLD58	0.27586	0.27161	1.00000	0.00000
8	HLD56	0.37931	0.48136	0.07269	0.00023
9	HLD118	0.49425	0.40695	0.06030	0.00024
10	HLD92	0.47126	0.49625	0.66928	0.00046
11	HLD93	0.45977	0.43054	0.62007	0.00049
12	HLD99	0.47126	0.45798	0.81724	0.00039
13	HLD88	0.47126	0.46462	1.00000	0.00000
14	HLD101	0.44828	0.5005	0.39246	0.00051
15	HLD67	0.36782	0.49963	0.01789	0.00013
16	HLD83	0.12644	0.11913	1.00000	0.00000
17	HLD114	0.3908	0.38004	1.00000	0.00000
18	HLD48	0.34483	0.35612	0.76566	0.00043
19	HLD124	0.42529	0.47631	0.36843	0.00049
20	HLD122	0.44828	0.39658	0.28068	0.00047
21	HLD125	0.27586	0.25566	0.68161	0.00048
22	HLD64	0.22989	0.23919	0.65599	0.00047
23	HLD81	0.32184	0.31626	1.00000	0.00000
24	HLD136	0.44828	0.48588	0.50757	0.00051
25	HLD133	0.51724	0.43485	0.08727	0.00029
26	HLD97	0.47126	0.47631	1.00000	0.00000
27	HLD40	0.5977	0.50282	0.08829	0.00027
28	HLD128	0.34483	0.40183	0.18898	0.00041
29	HLD39	0.43678	0.50123	0.28180	0.00045
30	HLD84	0.44828	0.42608	0.79945	0.00041

Adicionalmente podemos resaltar en este análisis que los valores de heterocigosidad en la mayoría de los marcadores están por encima de 0.4 y próximos a 0.5. Estos valores son importantes al tratarse de marcadores bialélicos, ya que para ser considerado un marcador bialélico altamente polimórfico sus dos alelos han de presentar valores de heterocigosidad próximos o iguales a 0.5. En estos análisis, nuestros resultados son coherentes con estos fundamentos. Sin embargo, de los treinta marcadores, cinco (HLD131, HLD58, HLD83, HLD125, HLD64) presentaron valores de heterocigosidad entre 0.12 y 0.30, es decir son poco informativos o polimórficos.

Por otro lado, podemos entender nuestros resultados de heterocigosidad como un modo intuitivo de medir la diversidad genética a través de la heterocigosidad (H), expresado como el porcentaje promedio de individuos heterocigotos por cada locus, en nuestro caso particular tenemos 30 locus distribuidos a lo largo de nuestro genoma en 19 cromosomas autosómicos (Investigator-DIPlex Kit, Qiagen-Germany).

También realizamos una comparación de los valores del Equilibrio de Hardy-Weingberg entre Perú, España Central, País Vasco y Uruguay. Cuando analizamos los 30 marcadores, no observamos desviaciones al Equilibrio de Hardy-Weingberg excepto en los siguientes marcadores por cada país (Tabla 11): Perú HLD45 (0.04595), HLD131 (0.03604), HLD67 (0.01728); España Central HLD81 (0.00804), HLD128 (0.01650); País Vasco HLD101 (0.03773). Sin embargo realizando la corrección de Bonferroni ($P\text{-value} = 0.001666$) se observó que los 30 marcadores de Inserción y Deleción no muestran desviaciones al Equilibrio de Hardy-Weingberg en ninguna de estas poblaciones. Estos resultados no necesariamente conducen a pensar que las poblaciones se encuentren en Equilibrio de Hardy-Weingberg, pero nos muestra las variaciones de la heterocigosidad en cada población, donde se observa que a pesar de no mostrar desviaciones al Equilibrio de Hardy-Weingberg en estas poblaciones, la variación de la heterocigosidad en cada marcador es diferente.

Tabla 11. Comparación del equilibrio de Hardy-Weingberg en cuatro poblaciones mediante 30 INDELS

N° Locus	LOCUS	PERU			ESPAÑA CENTRAL			PAIS VASCO			URUGUAY		
		Het.Obs.	Het. Esp	P-value	Het. Obs.	Het.Esp.	P-value	Het.Obs.	Het.Esp.	P-value	Het.Obs.	Het.Esp.	P-value
1	HLD77	0.52874	0.46774	0.25441	0.46479	0.48397	0.80575	0.48333	0.45364	0.77431	0.45038	0.47559	0.58132
2	HLD45	0.33333	0.42608	0.04595	0.45070	0.50105	0.47467	0.56667	0.50420	0.43637	0.51908	0.50048	0.72853
3	HLD131	0.27586	0.36835	0.03604	0.49296	0.50355	1.00000	0.40000	0.47675	0.27401	0.45802	0.50048	0.37978
4	HLD70	0.45977	0.45445	1.00000	0.46479	0.47797	1.00000	0.46667	0.45882	1.00000	0.38931	0.45032	0.12508
5	HLD6	0.48276	0.47891	1.00000	0.56338	0.50345	0.34773	0.51667	0.50070	1.00000	0.48092	0.48776	1.00000
6	HLD111	0.43678	0.49485	0.28287	0.49296	0.50355	1.00000	0.48333	0.49734	1.00000	0.45038	0.50145	0.29253
7	HLD58	0.27586	0.27161	1.00000	0.54930	0.49995	0.47452	0.40000	0.46835	0.27653	0.58779	0.49770	0.05188
8	HLD56	0.37931	0.48136	0.07156	0.40845	0.37409	0.53471	0.53333	0.47675	0.41770	0.51908	0.49346	0.59571
9	HLD118	0.49425	0.40695	0.06172	0.52113	0.48916	0.62785	0.51667	0.49286	0.79377	0.41985	0.47899	0.20071
10	HLD92	0.47126	0.49625	0.66774	0.52113	0.50315	0.81482	0.43333	0.49020	0.43013	0.44275	0.48644	0.36709
11	HLD93	0.45977	0.43054	0.61945	0.56338	0.50345	0.34808	0.43333	0.50364	0.30916	0.44275	0.48902	0.28673
12	HLD99	0.47126	0.45798	0.81858	0.49296	0.48397	1.00000	0.46667	0.46835	1.00000	0.46565	0.47197	1.00000
13	HLD88	0.47126	0.46462	1.00000	0.50704	0.49146	0.81095	0.46667	0.48403	0.79347	0.54198	0.48507	0.20638
14	HLD101	0.44828	0.50050	0.38756	0.54930	0.50315	0.48181	0.63333	0.49524	0.03773	0.48092	0.50145	0.72782
15	HLD67	0.36782	0.49963	0.01728	0.49296	0.47118	0.80059	0.50000	0.50196	1.00000	0.47328	0.49136	0.72335
16	HLD83	0.12644	0.11913	1.00000	0.45070	0.50265	0.47434	0.55000	0.49734	0.44080	0.51145	0.49618	0.85935
17	HLD114	0.39080	0.38004	1.00000	0.57746	0.50315	0.23700	0.38333	0.44244	0.37761	0.49618	0.48507	0.85794
18	HLD48	0.34483	0.35612	0.76545	0.54930	0.49715	0.47165	0.40000	0.50196	0.12721	0.43511	0.50004	0.16310
19	HLD124	0.42529	0.47631	0.36716	0.50704	0.45949	0.44084	0.43333	0.43641	1.00000	0.50382	0.46188	0.34290
20	HLD122	0.44828	0.39658	0.27963	0.49296	0.48397	1.00000	0.36667	0.44818	0.24241	0.45038	0.49244	0.37432
21	HLD125	0.27586	0.25566	0.68152	0.45070	0.50105	0.47673	0.58333	0.50294	0.29862	0.47328	0.49838	0.59946
22	HLD64	0.22989	0.23919	0.65728	0.40845	0.48397	0.21980	0.56667	0.50364	0.43851	0.43511	0.43453	1.00000
23	HLD81	0.32184	0.31626	1.00000	0.64789	0.49146	0.00804	0.43333	0.50364	0.30836	0.45038	0.49443	0.37411
24	HLD136	0.44828	0.48588	0.51127	0.52113	0.49715	0.81007	0.40000	0.42353	0.75862	0.54198	0.49443	0.29018
25	HLD133	0.51724	0.43485	0.08489	0.38028	0.47118	0.12786	0.50000	0.47675	0.78837	0.47328	0.49838	0.59977
26	HLD97	0.47126	0.47631	1.00000	0.50704	0.50265	1.00000	0.35000	0.48053	0.05440	0.48092	0.50145	0.72779
27	HLD40	0.59770	0.50282	0.08924	0.42254	0.48666	0.32703	0.50000	0.49524	1.00000	0.58015	0.50048	0.07886
28	HLD128	0.34483	0.40183	0.19051	0.64789	0.49865	0.01650	0.55000	0.50406	0.60513	0.41985	0.49618	0.11019
29	HLD39	0.43678	0.50123	0.27926	0.43662	0.48397	0.46217	0.41667	0.40210	1.00000	0.50382	0.48059	0.59151
30	HLD84	0.44828	0.42608	0.79948	0.45070	0.48107	0.62715	0.53333	0.49020	0.59611	0.54198	0.49770	0.37796

Como explicamos en párrafos anteriores, la estimación de la diversidad genética se ha considerado en este análisis como una medida de la heterocigosidad media. En ese sentido se observa que los valores de heterocigosidad media se encuentran entre 0.41 a 0.48 en las cuatro poblaciones (Tabla12) Perú, España Central, País Vasco y Uruguay, es decir, muestran una diversidad genética relativamente alta, ya que estos marcadores al ser bialélicos, sus máximos valores de heterocigosidad estarían alrededor de 0.5. Asimismo, se observa que la heterocigosidad media del

Perú es 0.41 mientras que en España Central, País Vasco y Uruguay está alrededor de 0.48, estos valores son coherentes con los resultados obtenidos en los análisis de estructura poblacional (Fig. 11), donde el Perú se muestra como un grupo poblacional diferente de España Central, País Vasco y Uruguay.

Tabla 12. Comparación de la diversidad genética con otras poblaciones

N° Locus	Perú	España central	País Vasco	Uruguay	Het.Medía	Desv.Est.	Het.Total
1	0.47406	0.48397	0.45364	0.47559	0.47182	0.01287	0.49759
2	0.45557	0.50105	0.50420	0.50048	0.49033	0.02323	0.49171
3	0.36247	0.50355	0.47675	0.50048	0.46081	0.06665	0.48222
4	0.46861	0.47797	0.45882	0.45032	0.46393	0.01198	0.46169
5	0.45398	0.50345	0.50070	0.48776	0.48647	0.02271	0.48514
6	0.48094	0.50355	0.49734	0.50145	0.49582	0.01025	0.49990
7	0.25697	0.49995	0.46835	0.49770	0.43074	0.11674	0.47779
8	0.45780	0.37409	0.47675	0.49346	0.45053	0.05300	0.46426
9	0.39755	0.48916	0.49286	0.47899	0.46464	0.04511	0.46169
10	0.50051	0.50315	0.49020	0.48644	0.49507	0.00802	0.49503
11	0.46470	0.50345	0.50364	0.48902	0.49020	0.01833	0.49036
12	0.44348	0.48397	0.46835	0.47197	0.46694	0.01701	0.49914
13	0.47187	0.49146	0.48403	0.48507	0.48311	0.00818	0.49696
14	0.50052	0.50315	0.49524	0.50145	0.50009	0.00341	0.50037
15	0.50227	0.47118	0.50196	0.49136	0.49169	0.01458	0.49706
16	0.13675	0.50265	0.49734	0.49618	0.40823	0.18101	0.45512
17	0.39416	0.50315	0.44244	0.48507	0.45620	0.04857	0.46239
18	0.32665	0.49715	0.50196	0.50004	0.45645	0.08656	0.46966
19	0.46267	0.45949	0.43641	0.46188	0.45511	0.01254	0.45662
20	0.39416	0.48397	0.44818	0.49244	0.45469	0.04468	0.45883
21	0.24126	0.50105	0.50294	0.49838	0.43591	0.12978	0.45737
22	0.20402	0.48397	0.50364	0.43453	0.40654	0.13811	0.41098
23	0.32179	0.49146	0.50364	0.49443	0.45283	0.08751	0.49731
24	0.48988	0.49715	0.42353	0.49443	0.47625	0.03527	0.48558
25	0.44443	0.47118	0.47675	0.49838	0.47269	0.02219	0.50060
26	0.48057	0.50265	0.48053	0.50145	0.49130	0.01242	0.50043
27	0.49856	0.48666	0.49524	0.50048	0.49524	0.00611	0.49972
28	0.42563	0.49865	0.50406	0.49618	0.48113	0.03715	0.48379
29	0.50117	0.48397	0.40210	0.48059	0.46696	0.04417	0.48362
30	0.43127	0.48107	0.49020	0.49770	0.47506	0.02997	0.50001
Media	0.41481	0.48791	0.47939	0.48812	0.46756	0.03540	0.48076
Desv.Est.	0.09646	0.02453	0.02706	0.01594	0.04100	0.03728	0.02098

7.4. Análisis de varianza molecular (AMOVA)

Del análisis de varianza molecular no se observó diferencias significativas entre los grupos Norte, Centro y Sur del Perú, tampoco se observó diferencias significativas entre las poblaciones dentro de los grupos ni entre los individuos dentro de las poblaciones. La única diferencia significativa resultó entre individuos (97.86 %). Este resultado es coherente con las características de los marcadores moleculares de uso forense, los cuales presentan muy baja variabilidad interpoblacional y alta variabilidad a nivel de individuos, esto para poder discriminar con alta probabilidad a dos individuos no emparentados dentro de una población. Asimismo, considerando que estamos comparando poblaciones muy próximas, se deduce que el flujo génico entre éstas es muy alta y por lo tanto los valores de diferenciación estarán próximos a cero, es así que en nuestro análisis los índices de fijación entre individuos dentro de las poblaciones fue 0.01903, entre las poblaciones dentro de los grupos fue -0.04029, entre grupos 0.01023, entre individuos 5.04044 (Tabla 13), evidenciando que la mayor variabilidad se encuentra entre individuos de la población total.

Tabla 13. Resultados del Análisis de Varianza Molecular (AMOVA) entre norte, centro y sur, obtenido mediante el programa Arlequin V.3.5.

Fuente de Variación	d.f.	Suma de cuadrados	Componentes de varianza	Porcentaje de variación
Entre grupos	2	17.265	0.05271 Va	1.02
Entre las poblaciones dentro de los grupos	7	28.951	-0.04029 Vb	-0.78
Entre los individuos dentro de las poblaciones	126	659.736	0.09778 Vc	1.90
Entre individuos	136	685.5	5.04044 Vd	97.86
Total	271	1391.452	5.15065	100.00

En la Tabla 14 se observa alta probabilidad de no diferenciación en Vd, Vc, Vb, Va, cuyos valores de significancia son superiores a 0.005. Asimismo, en los análisis de *Fst* (Cuadro 5 y 6) no se observa diferencias significativas entre las regiones del norte, centro y sur. Adicionalmente se tiene resultados congruentes de no diferenciación en los análisis de estructura poblacional (Fig.10).

Tabla 14.Significancia del Análisis de Varianza Molecular

Test de Significancia (10100 permutaciones)		
Vd and FIT :	P(rand. value < obs. value)=	0.22297
	P(rand. value = obs. value)=	0.00921
	P-value =	0.23218
Vc and FIS :	P(rand. value > obs. Value =	0.21475
	P(rand. value = obs. value)=	0.00891
	P-value =	0.22366
Vb and FSC :	P(rand. value > obs. Value =	0.97119
	P(rand. value = obs. value)=	0.00000
	P-value =	0.97119
Va and FCT :	P(rand. value > obs. value)=	0.00149
	P(rand. value = obs. value)=	0.00079
	P-value =	0.00228

7.5. Análisis de diferenciación poblacional (*Fst*)

Cuadro 5. Muestra la correlación de la significancia (P-value) con las distancias genéticas (Fst) entre las poblaciones de Perú, España, País Vasco y Uruguay.

	Perú Norte	Perú Centro	Perú Sur	España Central	País Vasco	Uruguay Noreste	Uruguay Central	Uruguay Litoral	Uruguay Sur
Perú-Norte	*	0.36913	0.01393	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Perú-Centro	0.00153	*	0.25461	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Perú-Sur	0.01896	0.00400	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
España-Central	0.06857	0.10026	0.09952	*	0.00999	0.18778	0.01978	0.22040	0.20057
País-Vasco	0.07630	0.10763	0.11259	0.00529	*	0.03096	0.04050	0.05179	0.00005
Uruguay-Noreste	0.04964	0.07638	0.08908	0.00324	0.00889	*	0.74305	0.98627	0.67542
Uruguay-Central	0.06072	0.08697	0.10769	0.01640	0.01446	-0.00686	*	0.06772	0.48454
Uruguay-Litoral	0.05151	0.08430	0.09810	0.00216	0.00692	-0.01262	0.01179	*	0.71673
Uruguay-Sur	0.04845	0.07576	0.08365	0.00152	0.01108	-0.00153	0.00104	-0.00229	*

Las distancias genéticas (*Fst*) comparadas entre las regiones del norte, centro, y sur del Perú (Cuadro 5) presentan valores próximos a cero, es decir, podemos inferir que existe un alto flujo génico entre la población mestiza peruana, además, estas distancias geográficamente son cortas por lo que es coherente con los valores de *Fst*. Asimismo, muestra valores superiores al nivel de significancia (*P-Value* = 0.05), lo que permite sustentar que no existe diferencias significativas entre las tres regiones analizadas. Sin embargo, si observamos los niveles de significancia (Cuadro 5 y 6) entre Perú y España Central, País Vasco y Uruguay presentan valores de cero (*P-Value* <0.05), de estos resultados podemos inferir que a una escala intercontinental se puede apreciar diferencias significativas entre las poblaciones analizadas.

Cuadro 6. Comparación de valores de F_{st} y niveles de significancia entre las poblaciones de Perú, España Central, País Vasco y Uruguay.

F_{ST}	P-value	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	*		0.83294	0.26240	0.92993	0.47890	0.44764	0.28148	0.90006	0.20262	0.00709	0.00000	0.00000	0.00000	0.00075	0.00000	0.00000
2	-0.0084	*		0.89332	0.80947	0.15962	0.74340	0.81596	0.90786	0.81217	0.04085	0.00000	0.00000	0.00005	0.00070	0.00000	0.00000
3	0.00698	-0.01369	*		0.66548	0.54587	0.75448	0.69940	0.68796	0.25636	0.00869	0.00000	0.00000	0.00000	0.00035	0.00000	0.00000
4	-0.01275	-0.00984	-0.00498	*		0.34196	0.72237	0.37737	0.53493	0.38805	0.07456	0.00000	0.00000	0.00030	0.00085	0.00015	0.00000
5	0.00146	0.01251	-0.00087	0.00539	*		0.43045	0.40483	0.50022	0.26669	0.04999	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
6	0.00224	-0.0065	-0.00708	-0.00584	0.00232	*		0.55077	0.72142	0.36888	0.03925	0.00000	0.00000	0.00000	0.00050	0.00000	0.00000
7	0.00549	-0.00907	-0.00595	0.00284	0.00158	-0.00185	*		0.62373	0.78240	0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
8	-0.01094	-0.01262	-0.00486	0.00105	0.00162	-0.0052	-0.00302	*		0.27858	0.07062	0.00000	0.00000	0.00005	0.00025	0.00000	0.00000
9	0.01279	-0.01142	0.0084	0.00522	0.00951	0.00502	-0.01092	0.01137	*		0.82185	0.00000	0.00000	0.00000	0.00005	0.00000	0.00000
10	0.02796	0.02111	0.02856	0.01664	0.01732	0.01923	0.02037	0.01725	-0.01435	*		0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
11	0.05734	0.06345	0.07269	0.0556	0.09318	0.07436	0.08739	0.0771	0.09785	0.08715	*		0.04705	0.21111	0.02118	0.27244	0.1553
12	0.0671	0.07905	0.09046	0.06981	0.10663	0.09319	0.11221	0.09197	0.12085	0.10218	0.00370	*		0.01813	0.03366	0.05219	0.0002
13	0.05071	0.05036	0.06016	0.04384	0.07881	0.05875	0.07427	0.05928	0.09269	0.08570	0.00309	0.01053	*		0.74889	0.97743	0.71648
14	0.05411	0.05789	0.06632	0.06532	0.07932	0.06223	0.08070	0.05531	0.10851	0.10463	0.01691	0.01562	-0.00724	*		0.09434	0.45398
15	0.04658	0.05045	0.06233	0.04665	0.08113	0.07415	0.07848	0.06895	0.09293	0.09760	0.00170	0.00710	-0.01228	0.01007	*		0.82166
16	0.03654	0.04104	0.04667	0.04001	0.06460	0.05151	0.06164	0.05083	0.07376	0.07469	0.00196	0.01018	-0.00209	0.00148	-0.00371	*	

Los valores de F_{st} comparados corresponden a las siguientes poblaciones:

- | | | |
|---------------------------|--------------------------|---------------------|
| 1. Piura-Tumbes | 07. Huánuco-Ucayali. | 13. Uruguay-noreste |
| 2. Amazonas-San Martín | 08. Junín-Cerro de Pasco | 14. Uruguay central |
| 3. Lambayeque-La Libertad | 09. Ayacucho-Cuzco | 15. Uruguay-litoral |
| 4. Cajamarca | 10. Puno-Arequipa | 16. Uruguay-sur |
| 5. Ancash-norte de Lima | 11. España-central | |

7.6. Estimación del Desequilibrio de Ligamiento

Los 30 marcadores analizados muestran evidencias suficientes para no descartar la posibilidad de que se encuentran en desequilibrio de ligamiento (Tabla 15) pues casi en su totalidad los marcadores presentan valores por debajo del nivel de significancia ($P\text{-value} = 0.05$), es decir, segregan de manera independiente y no son afectados por la recombinación genética, ni por mecanismos de selección natural, podemos decir que estos marcadores son estadísticamente independientes.

Tabla 15. Tabla de significancia del Desequilibrio de Ligamiento (Nivel de significancia 0.05)

#	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
0	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-
26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-
27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-
29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*

7.7. Inferencia de historia evolutiva

En las Tablas 20 y 21 basados en los análisis de distancias genéticas, en general se observa un resultado congruente con los antecedentes históricos de Perú, en el cual, las poblaciones peruanas se vinculan más entre sí, comparado con la población de España (colonizadores a inicios del Siglo XVII), aunque a nivel intrapoblacional en el Perú se observa agrupamientos entre norte, centro y sur, principalmente se diferencia el Sur (Ayacucho, Cuzco, Puno y Arequipa) del resto de poblaciones, posiblemente por su alto componente andino. Sin embargo, ello estadísticamente no resultó ser significativo como para evidenciar subestructura poblacional en nuestro país, esto podría atribuirse al tamaño muestral analizado (136 peruanos mestizos del total de 32 millones de peruanos habitantes hasta el año 2018), también podría deberse al tipo de marcadores empleados (INDELs) los cuales son estandarizados para uso forense. Sin embargo, además de los objetivos principales de este estudio estos análisis nos permitieron reflejar cierta tendencia de agrupamientos coherentes con la historia y con nuestra diversidad genética y geográfica, resultados susceptibles de complementar más adelante con otros sistemas de marcadores y con mayor tamaño muestral, con los cuales se pueda confirmar o refutar nuestros resultados.

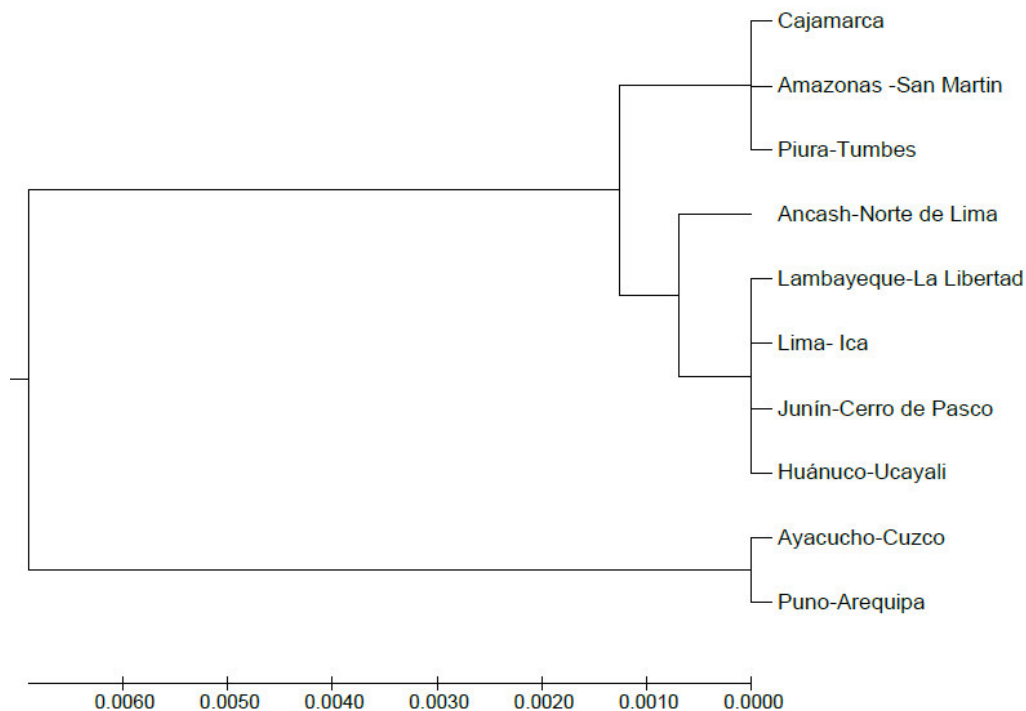


Figura 12. Arbol de distancias genéticas de UPGMA, de la población peruana, muestra el agrupamiento en tres regiones: Norte (Cajamarca, Amazonas, San Martín, Piura y Tumbes), Centro (Ancash, norte de Lima, Lima, Ica, Junín, Cerro de Pasco, Huánuco y Ucayali), Sur (Ayacucho, Cuzco, Puno y Arequipa).

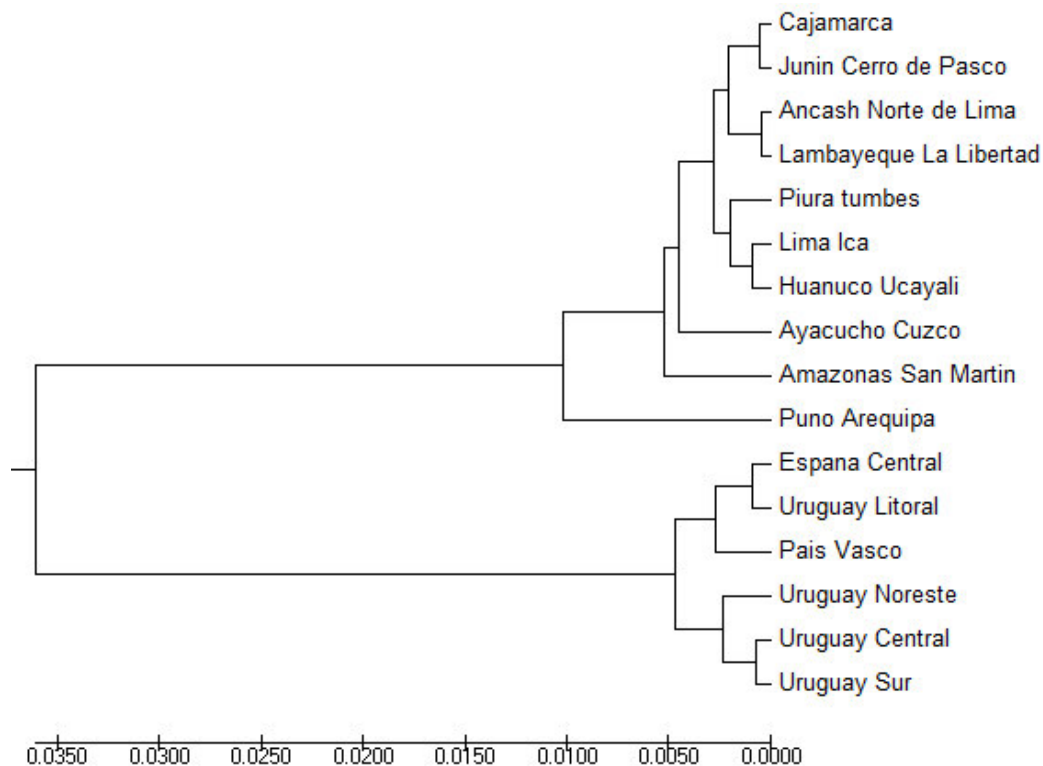


Figura 13. Árbol de distancias Genéticas de UPGMA la población peruana y España Central, País Vasco y Uruguay se puede observar principalmente dos agrupamientos entre Perú y los otros países estudiados.

8. ANALISIS DE PARAMETROS ESTADÍSTICOS DE INTERÉS FORENSE

Para poder analizar los parámetros estadísticos de interés forense, previamente analizamos si los 30 marcadores INDELs presentan o no desviaciones al Equilibrio de Hardy–Weinberg (EHW). Como se describió anteriormente, los 30 marcadores no presentaron desviaciones al equilibrio de Hardy–Weinberg, esto nos permitió analizar los INDELs aplicando la fórmula del binomio al cuadrado de tal manera que estimamos las frecuencias genotípicas a partir de las frecuencias alélicas. Con los datos de frecuencias genotípicas estimamos los parámetros estadísticos de interés forense tales como: Parámetros de paternidad (Poder de exclusión, Índice típico de paternidad), contenido de información polimórfica (PIC), Probabilidad de Match (PM) y Poder de Discriminación (PD), (Cuadro 7).

Cuadro 7. Parámetros estadísticos de interés forense: Parámetros de paternidad, contenido de Información Polimórfica (PIC), Probabilidad de Match (PM) y Poder de Discriminación (PD).

PARÁMETROS ESTADÍSTICOS DE INTERÉS FORENSE											
N°	LOCUS	Frecuencias Alélicas		Frecuencias Genotípicas			Parametros de paternidad		PIC	Probabilidad de Match (PM)	Poder de Discriminación (PD)
		p	q	p ²	2pq	q ²	Poder de Exclusión	Indice Típico de Paternidad			
01	HLD77	36.78%	63.22%	0.135288678	0.46505483	0.39965649	0.214	1.06	0.36	0.39430433	0.60569567
02	HLD45	30.46%	69.54%	0.09277976	0.42363588	0.48358436	0.078	0.75	0.33	0.42192928	0.57807072
03	HLD131	75.86%	24.14%	0.575505351	0.36623068	0.05826397	0.054	0.69	0.30	0.46872601	0.53127399
04	HLD70	34.48%	65.52%	0.118906064	0.45184304	0.42925089	0.155	0.93	0.35	0.40255712	0.59744288
05	HLD6	39.08%	60.92%	0.152728234	0.47615273	0.37111904	0.173	0.97	0.36	0.38777667	0.61222333
06	HLD111	56.32%	43.68%	0.317214956	0.49200687	0.19077817	0.138	0.89	0.37	0.37909240	0.62090760
07	HLD58	83.91%	16.09%	0.704056018	0.27004888	0.02589510	0.054	0.69	0.23	0.56929183	0.43070817
08	HLD56	39.66%	60.34%	0.15725327	0.47859691	0.36414982	0.102	0.81	0.36	0.38638868	0.61361132
09	HLD118	71.84%	28.16%	0.516085348	0.40461091	0.07930374	0.183	0.99	0.32	0.43634316	0.56365684
10	HLD92	55.75%	44.25%	0.310774211	0.49339411	0.19583168	0.164	0.95	0.37	0.37836840	0.62163160
11	HLD93	31.03%	68.97%	0.096313912	0.42806183	0.47562426	0.155	0.93	0.34	0.41873173	0.58126827
12	HLD99	64.94%	35.06%	0.421753204	0.45534417	0.12290263	0.164	0.95	0.35	0.40031913	0.59968087
13	HLD88	36.21%	63.79%	0.131093936	0.46195006	0.40695600	0.164	0.95	0.36	0.39619667	0.60380333
14	HLD101	53.45%	46.55%	0.285671819	0.49762188	0.21670630	0.146	0.91	0.37	0.37619754	0.62380246
15	HLD67	54.02%	45.98%	0.291848329	0.49676311	0.21138856	0.096	0.79	0.37	0.37663416	0.62336584
16	HLD83	93.68%	6.32%	0.877559783	0.11844365	0.00399656	0.013	0.57	0.11	0.78415604	0.21584396
17	HLD114	74.71%	25.29%	0.558197913	0.37785705	0.06394504	0.108	0.82	0.31	0.45844983	0.54155017
18	HLD48	22.99%	77.01%	0.05284714	0.35407584	0.59307702	0.084	0.76	0.29	0.47990287	0.52009713
19	HLD124	38.51%	61.49%	0.148269256	0.47357643	0.37815431	0.130	0.87	0.36	0.38925909	0.61074091
20	HLD122	72.99%	27.01%	0.532732197	0.39430572	0.07296208	0.146	0.91	0.32	0.44460406	0.55539594
21	HLD125	85.06%	14.94%	0.723477342	0.25419474	0.02232792	0.054	0.69	0.22	0.58853297	0.41146703
22	HLD64	13.79%	86.21%	0.01902497	0.23781213	0.74316290	0.038	0.65	0.21	0.60920766	0.39079234
23	HLD81	19.54%	80.46%	0.038182058	0.31444048	0.64737746	0.073	0.74	0.27	0.51942826	0.48057174
24	HLD136	40.80%	59.20%	0.166501519	0.48308892	0.35040957	0.146	0.91	0.37	0.38388452	0.61611548
25	HLD133	68.39%	31.61%	0.467730215	0.43235566	0.09991412	0.203	1.04	0.34	0.41568580	0.58431420
26	HLD97	61.49%	38.51%	0.378154314	0.47357643	0.14826926	0.164	0.95	0.36	0.38925909	0.61074091
27	HLD40	49.43%	50.57%	0.244285903	0.49993394	0.25578016	0.288	1.24	0.37	0.37503304	0.62496696
28	HLD128	72.41%	27.59%	0.524375743	0.39952438	0.07609988	0.084	0.76	0.32	0.44038084	0.55961916
29	HLD39	52.87%	47.13%	0.279561369	0.49834853	0.22209010	0.138	0.89	0.37	0.37582983	0.62417017
30	HLD84	69.54%	30.46%	0.483584357	0.42363588	0.09277976	0.146	0.91	0.33	0.42192928	0.57807072
Se puede observar que los 30 marcadores INDELS, muestran un alto Poder de Discriminación (PD), y el valor de Probabilidad de Match acumulado (PM) es bajo, siendo estas características lo deseado en marcadores forenses, es decir valores bajos de PM y bastante altos de PD.										PM-Total	PD-Total
										0.0000000001426561820	0.9999999999857340

El poder discriminación indica que dos individuos no relacionados tomados al azar puedan distinguirse genéticamente en uno o más marcadores, el valor de este parámetro en el set de 30 marcadores analizados fue de 0.9999999999857340, mostrando un alto poder de discriminación (PD). Asimismo, la Probabilidad de Coincidencia (PM) indica que dos individuos no relacionados tomados al azar de la misma población puedan coincidir en su genotipo con determinado sistema de marcadores, normalmente se espera que los valores de PM sean muy bajos. En nuestros análisis la probabilidad de coincidencia acumulada fue de 0.00000000001426561820. En marcadores forenses el Poder de Discriminación (PD) y la Probabilidad de Coincidencia (PM) son conceptos opuestos, es decir, se espera que para calificar a un sistema o set de marcadores como válidos o adecuados deben tener un alto Poder de Discriminación y muy baja Probabilidad de Coincidencia, los cuales son congruentes con nuestros resultados.

Con respecto al Índice de Contenido Polimórfico (PIC), los valores en promedio resultaron alrededor de 0.32, con excepción del marcador HLD83 (0.11) y HLD58 (0.23). En general se puede apreciar que los 30 marcadores analizados son marcadores medianamente informativos con respecto a su polimorfismo según la clasificación de Botstein et al., (1980), ya que un marcador bialélico con alto polimorfismo, se esperaría que sus valores se encuentren alrededor de 0.5. También analizamos el poder de exclusión e índice típico de paternidad para cada marcador como se muestra en el Cuadro 7 los cuales son de mucha importancia en los análisis de cálculos probabilísticos para incluir o excluir imputados en una prueba de paternidad.

9. DISCUSIÓN

Los análisis de estructuración poblacional empleando 30 marcadores INDELs en el Perú revelan una población homogénea, donde no se distinguen diferencias significativas entre el Norte, Centro y Sur del Perú. Considerando nuestros análisis en peruanos mestizos, estos resultados fueron similares a los obtenidos en otras poblaciones (España y Uruguay) con el mismo conjunto de marcadores (Saís, 2012) no distinguiéndose diferencias entre grupos de un mismo territorio o país. Del mismo modo, estudios de diferenciación poblacional entre europeos no mostraron diferencias significativas para estos marcadores, sin embargo, se observó diferencias significativas entre poblaciones de diferentes continentes (Martín, et al 2012). En nuestros análisis también observamos agrupamiento o diferenciación al comparar perfiles genéticos entre la población de España, País Vasco, Uruguay y Perú. De estos análisis se puede inferir que a nivel intercontinental este sistema permite distinguir genéticamente una población de otra mas no entre grupos de una misma población.

La consistencia de estos resultados se corroboró mediante el programa Structure V.2.3.4 (Pritchard et al., 2009) con el cual se observó que la población mestiza del Perú no presenta estructuración entre el Norte, Centro y Sur del Perú. Sin embargo, si analizamos la estructura poblacional agregando genotipos del mismo conjunto de marcadores de poblaciones europeas como el de España, País Vasco, o de poblaciones geográficamente cercanas como Uruguay, en estas condiciones el sistema INDELs muestra un comportamiento diferente, pues permite distinguir genéticamente la población peruana de estos países formando visiblemente dos grupos ($K = 2$). Así observamos que a pesar de presentar el antecedente histórico de colonización por parte de españoles, la población peruana muestra básicamente un componente genético propio que se distingue de los españoles, del País Vasco y también de Uruguay. En nuestros análisis también se puede observar que la población uruguaya posee mayor similitud genética con la población española a diferencia del Perú, es decir el acervo genético de Uruguay posee un alto componente europeo. En nuestro caso luego de observar estos resultados podemos respaldar la hipótesis de que a pesar de poseer históricamente poblaciones

europeas, asiáticas y africanas en nuestra población genéticamente poseemos un mayor componente de americanos nativos ya sean andinos o amazónicos y en menor proporción un componente extranjero.

Por otro lado, los análisis de Varianza Molecular (AMOVA) y distancias genéticas (F_{st}) mostraron valores significativos de no diferenciación, esto es coherente con los resultados de estructuración poblacional, se puede observar que existe alta diversidad genética (97.86%) entre individuos de la misma población, mas no entre grupos poblacionales. Asimismo, Martín y colaboradores (2013) empleando el mismo conjunto de marcadores en poblaciones de España Central y el País Vasco observaron que en los análisis de AMOVA se distingue diferencias significativas dentro de las poblaciones mas no entre poblaciones.

Con respecto al análisis de distancias genéticas de a pares (F_{st}) se observó que los valores de F_{st} , en la población peruana resultaron negativos o próximos a cero (Cuadro 6), y los valores de significancia (P-value) mayores a 0.05, es decir, no refleja diferencias genéticas significativas entre los grupos de nuestra población. Por otro lado, si analizamos a una escala intercontinental las distancias genéticas con España y el País Vasco en comparación con el Perú y Uruguay, los valores de F_{st} muestran diferencias significativas entre estos países (P-value < 0.05), es decir, a pesar del antecedente histórico que poseemos por la colonización de España, podemos observar evidencias de estructuración poblacional entre España y el Perú. En consecuencia, cada país posee su propio patrón de flujo génico.

Cabe resaltar que en el caso de Uruguay comparado con España y el País Vasco los valores de F_{st} , son similares, es decir Uruguay evidencia un mayor componente genético europeo español a diferencia de Perú. De estas observaciones, respecto de los marcadores INDELs, podemos inferir que a nivel de distancias geográficas cortas dentro del Perú no podemos observar diferencias genéticas significativas. Sin embargo ocurre algo diferente si empleamos estos mismos marcadores a una escala intercontinental como ya se discutió anteriormente.

Con respecto al análisis del Equilibrio de Hardy–Weinberg (H-W) en general no se observó desviaciones al Equilibrio de Hardy–Weinberg para los 30 marcadores INDELs estudiados.

Sobre los parámetros estadísticos de interés forense, el Poder de Discriminación (PD) acumulado ha sido alto (0.999999999857340), significa que en más de un 99.99999999% de las veces tenemos la probabilidad de excluir a un inocente en una prueba de ADN, dicho de otro modo, este sistema tiene un alto poder de discriminar un perfil genético de otro en un conjunto de evidencias que procedan de la misma escena de crimen o de determinado caso legal como pruebas de paternidad u otros casos relacionados.

Sobre la probabilidad de match (PM). Esta probabilidad en la práctica debería ser muy baja (muy cercano a cero), la probabilidad de match acumulada del set de 30 marcadores fue de 0.00000000001426561820, significa que al emplear este sistema en un caso forense existe una probabilidad muy baja (muy próximo a cero) de que dos individuos no relacionados compartan el mismo perfil genético.

También analizamos el Índice de Contenido Polimórfico (PIC) y se observó que los marcadores son medianamente informativos pues sus valores estuvieron alrededor de 0.32 en promedio, sin embargo los marcadores más polimórficos fueron los siguientes: HLD111, HLD92, HLD101, HLD67, HLD136, HLD39, HLD40, cuyos valores fueron de 0.37, mientras los marcadores menos informativos fueron: HLD81, HLD64, HLD125, HLD83, HLD58. Estas evidencias sugieren que en la investigación forense o investigación genético poblacional se debe considerar los marcadores más informativos.

Los análisis del árbol de distancias genéticas, si bien es cierto muestran un patrón de distribución entre el Norte, Centro y Sur del Perú, en el cual Puno-Arequipa y Ayacucho-Cuzco se diferencian aparentemente como un grupo de las otras regiones del país, sin embargo ello no resultó estadísticamente significativo, es decir la población mestiza del Perú se comporta como una unidad, lo cual también se pudo corroborar con los análisis de AMOVA y Distancias genéticas Fst. También se pudo observar en el análisis que la población peruana se distingue de la población de española, ello es coherente con los resultados anteriores donde los peruanos poseemos principalmente un componente genético americano nativo y muy poco de hispanos o de otras poblaciones.

10. CONCLUSIONES

- 10.1. La diversidad genética de la población mestiza estudiada mediante 30 marcadores estándar de inserción y delección no muestra diferencias significativas entre las regiones del norte, centro y sur.
- 10.2. A escala intercontinental, el conjunto de marcadores utilizados distingue genéticamente a la población peruana de las poblaciones de España, País Vasco y Uruguay formando visiblemente dos grupos.
- 10.3. La población peruana mestiza a pesar de poseer un acervo genético europeo, asiático, africano y de otras poblaciones debido a la migración y/o colonización, genéticamente posee un componente mayoritario de americanos nativos y en menor proporción un componente genético extranjero.
- 10.4. La mayor diversidad genética de la población mestiza evaluada con INDELs se centra a nivel de individuos (alta diversidad genética individual), mas no a nivel de grupos poblacionales.

- 10.5. Los 30 marcadores INDELs no mostraron desviaciones al Equilibrio de Hardy–Weinberg.
- 10.6. Los parámetros estadísticos de interés forense mostraron un alto Poder de Discriminación (99.99999999%) y baja probabilidad de coincidencia (0.0000000001426561820) lo que hace que este sistema sea una herramienta muy útil en aplicación de la genética forense.

11. RECOMENDACIONES

En concordancia con nuestro principal objetivo, caracterizamos genéticamente 136 personas mestizas de las regiones del Norte, Centro y Sur del Perú. Los resultados que obtuvimos sobre la heterosigosidad, diversidad genética, patrones de agrupamiento y demás variables discutidos en este estudio pueden ser complementados y/o fortalecidos siguiendo las siguientes recomendaciones:

- 11.1. Incrementando el número de muestras de las diferentes regiones del país que representen al Perú, considerando que a lo largo de todo el territorio peruano existen diversos grupos culturales y étnicos. Asimismo se pueden analizar empleando otros marcadores (STR, SNPs, INDELs, DNA mitocondrial), siguiendo las recomendaciones internacionales sobre la cantidad de muestras a analizar en genética de poblaciones.
- 11.2. En este caso particular empleamos un set de 30 marcadores de Inserción y Delección INDELs, estandarizados para uso en genética forense, es decir, la característica inherente en estos marcadores es su alta diversidad genética a nivel de individuos de una población mas no entre grupos poblacionales. Sin embargo, se puede emplear este sistema para observar algunas tendencias de agrupamiento entre poblaciones geográficamente distantes. Si se tuviese los objetivos de evaluar ancestría, genética de poblaciones y migración, en esos casos recomendamos emplear otros tipos de marcadores, como el ADN mitocondrial, STRs de cromosoma Y, STR de cromosoma X, INDELs de ancestría (que tambien existen) o los SNPs.
- 11.3. Otro aspecto muy importante a considerar es el genotipado, para ello el investigador debe controlar y estandarizar las etapas claves en la obtención de los perfiles genéticos para garantizar la robustez y la reproducibilidad de los datos ya que a partir de estos llegaremos a un conjunto de conclusiones. Así por ejemplo, deberá tener en cuenta lo siguiente: La cuantificación del ADN antes de amplificar por

Reacción en Cadena de la Polimerasa (PCR), se debe conocer la concentración, la pureza, los índices de degradación del material genético con que se va a trabajar, luego de obtener estos datos se debe uniformizar las concentraciones a una cantidad específica conforme a las recomendaciones del conjunto marcadores que se va a emplear, esto se debe realizar para cualquier tipo de marcadores usados en genética de poblaciones o genética forense.

- 11.4. Se recomienda trabajar, en la medida de lo posible, solamente con perfiles genéticos completos, es decir, para este tipo de estudios es recomendable trabajar con ADN no degradado, de preferencia que no sea de larga data. Esto evitará la ausencia de alelos en cada genotipo, permitirá contabilizar el total de alelos, y distinguiremos si realmente existe o no alelos nulos y desbalances en los picos alélicos, es decir, la calidad de los genotipos es un aspecto crítico a considerar en este tipo de estudios.
- 11.5. Por otro lado, es preciso resaltar que en la práctica forense no es recomendable emplear bases de datos poblacionales de otras poblaciones, sino más bien, una base de datos de ADN propio de cada población que refleje realmente sus características genéticas y que su uso esté basado en la genética de poblaciones del lugar donde se está aplicando la genética forense.
- 11.6. Finalmente si deseamos conocer la contribución genética de otros grupos en la población que estamos analizando recomendamos emplear muestras de referencia para la población que vamos a estudiar, por ejemplo, grupos étnicos o grupos de otros continentes que históricamente exista o se conozca el antecedente de su contribución genética. A partir de ello se medirá la variabilidad genética.

12. REFERENCIAS BIBLIOGRÁFICAS

Allendorf F.W. & G.K. Luikart. (2007). Conservation and the Genetics of Populations, 1st ed. Blackwell Publishing, Oxford.

Arce J. (2008). Grandes incendios urbanos: Mesa Redonda Lima 2001. Rev Perú Med Exp Salud Pública. 25(1): 118-124.

Bastos-Rodrigues L., Pimenta J.R. & S.D Pena. (2006) The genetic structure of human population studied through short insertion-deletion polymorphisms, Ann. Hum. Genet. 70, 658-665.

Butler J.M., Shen Y. & B.R. Mc.Cord. (2003). The development of reduced size STR amplicons as tools for the analysis of degraded DNA. J. Forensic Sci. 48:1054–1064.

Brenner C. & J. Morris. (1990). Paternity index calculations in single locus hypervariable DNA probes: validation and other studies, p. 21-53. In Proceedings for the International Symposium on Human Identification 1989. Promega Corporation, Madison, WI.

Butler J.M. (2005). Forensic DNA Typing 2nd Ed. Elsevier Academic Press, Burlington, London.

Bridges C.B. (1936). The bar “gene” a duplication. Science ;83(2148):210-1.

Berger J., Suzuki T., Senti K.A., Stubbs J., Schaffner G. & B.J. Dickson. (2001). Genetic mapping with SNP markers in Drosophila. Nat Genet 29:475-481.

Botstein D., White R.L., Skolnick M. & R.W Davis. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331.58.

Cabrero J. & J.P.M. Camacho. (2011) .Fundamentos de Genética de Poblaciones. Departamento de Genética, Facultad de Ciencias. Universidad de Granada. 18071-Granada.

Carvalho A, Cainé L, Carvalho, & M.F. Pinheiro. (2011). Application of indels (investigator DIPplex) in mixture samples. Forensic Science International: Genetics Supplement Series. 3(1):e351.e352.

- Castagnino J.M. (2000), Electroforesis capilar. *Bioquímica*, vol. 25 (1): 13-32.
- Casalla R.H.(2003). Medición de distancias genéticas en poblaciones de camarón blanco *Litopenaeus vannamei* en la costa ecuatoriana. Escuela Superior Politécnica del Litoral. Facultad de Ingeniería Marítima y Ciencias del Mar. Guayaquil-Ecuador.
- Consortium I.H.G.S (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.
- Comas, D., Calafell, F., Bendukidze, N., Fananas, L. & J. Bertranpetit. (2000). Georgian and kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am J Phys Anthropol*, 112, 5-16.
- Collins, F.S., Drumm, M.L., Cole, J.L., Lockwood, W.K., Vande Woude, G.F., & M.C Iannuzzi. (1987). Construction of a general human chromosome jumping library, with application to cystic fibrosis. *Science* 235: 1046–1049.
- Dawson E. Chen Y. Hunt S. Smink L.J, Hunt A. Rice K. Livingston S. Bumpstead S. Bruskiewich R. Cham P, et al. (2001) A SNP resource for human chromosome 22: Extracting dense clusters of SNPs from the genomic sequence. *Genome Res.* 11: 170-178.
- Excoffier L., Lischer H.E. (2010. *Mol Ecol Resour*); 10(3):564-7.
- Ellegreen H (2000), Microsatellite mutations in the germline: Implications for evolutionary inference. *Trends Genet* 16: 551-558.
- Evanno G., Regnaut, S. & J. Goudet. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, 14:2611–2620.
- Francez P.A.C., Ribeiro-Rodrigues E.M. & S.E Batista dos Santos. (2012). Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon, *Forensic Science International: Genetics* 6 (2012) 132–135.

Foissac A. Crouau-Roy B. Fauré S. Thomsen M. & A Cambon-Thomsen. (1997). Review article Microsatellites in the HLA region: an overview. *Tissue Antigens* 49: 197–214.

Friis S.L., Borsting C., Rockenbauer E., Poulsen L., Fredslund S.F., Tomas C. & N, Morling. (2012). Typing of 30 insertion/deletions in Danes using the first commercial indel kit Mentype1 DIPplex, *Forensic Science International: Genetics* 6 (2012) e72–e74.

González-Andrade F. (2006). Análisis Molecular de variación de polimorfismos STR autosómicos y de cromosoma Y, en grupos étnicos de Ecuador con aplicación medico-forense. Memoria de Tesis doctoral en Medicina. España: Universidad de Zaragoza.

González E.A. (2011). Análisis de la diversidad genética en poblaciones naturales de especies vegetales amenazadas: *Ilex perado* ssp. *Lopezlilloi* (aquifoliaceae), *Silene nocteolens* (Caryophyllaceae) y *Sorbus aria* (Rosaceae).

Hollard C., Mendisco F., Keyser C., Crubezy E., Ludes B. (2011). First application of the Investigator DIPplex indels typing kit for the analysis of ancient DNA samples *Forensic Sci. Int. Genet. Suppl. Series 3*- 393–394.

Investigator DIPplex Handbook. (2013). For multiplex amplification of 30 deletion/insertion polymorphisms, plus Amelogenin. Sample & Assay Technologies. www.qiagen.com.

Jamieson A., (1994). The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Genetics* 25: 37-44.

Jones D.A, (1972). Blood lood samples- probalility of discrimination. *J Forensic Sci Soc* 12: 355-359.

Krenke B., Tereba A., Anderson S., Buel E., Culhane S., Finis C., Christine S. Tomsey, Zachetti J.M., Masibay A., Rabbach D.R., Amiott E.A. & C.J. Sprecher. (2002). Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 47: 773-785.

Lee H.S., Lee J.W., Han G.R. & J.J Hwang. (1998). Statistical parameters for evaluating the identification power of STR Loci. *Korean J Leg Med* 22: 13-19.

Li C., Zhao S., Zhang S., Li L., Liu Y., Chen J. & J. Xue. (2011). Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population, *Forensic Science International: Genetics* 5 (2011) e27–e30.

Lupski J.R., Roth J.R. & G.M Weinstock. (1996). Chromosomal duplications in bacteria, fruit flies, and humans, *Am. J. Hum. Genet.* 58:21-27.

Mills R.E., Lutting C.T., Larkins C.E, Beauchamp A., Tsui C., Pittard W.S. & S.E. Devine. (2006). And initial map of insertion and deletion (INDEL) variation in the huma genome, *Genome Res.* 16, 1182-1190.

Nachman M.W. & S.L Crowell (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297-304.

Nei M. & W. H. Li. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc.Natl.Acad.Sci.USA* 76:5269-5273.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.

Ostertag, E.M. & H.H Kazazian. (2001). Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35: 501–538.

Martín P., Garcia O., Heinrichs B., Yurrebaso I., Aguirre A. & A. Alonso. (2012). Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations. *Forensic Science International : Genetics* 7 (2013) e27-e30. <https://doi.org/10.1016/j.fsigen.2012.10.003>.

Pereira R., Phillips C., Alves C., Amorim A., Carracedo A. & L. Gusmão. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis* 2009, 30, 3682–3690.

Pereira R. & L. Gusmão. (2012). Capillary electrophoresis of 38 noncoding biallelic mini-indels for degraded samples and as complementary tool in paternity testing. *Methods Mol Biol* 830: 141-157.

Pimenta J.R. & S.D.J. Pena. (2010). Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. *Genet. Mol. Res.* 9 (1): 601-607.

Pritchard K., Wen X. & D Falush. (2009). Documentation for structure software: Version 2.3. 4.

Reynolds J. Weir B.S. & C.C Cockerham. (1983). Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.

Romanini C., Catelli M.L., Borosky A., Pereira R., Romero M., Salado Puerto M., Phillips C., Fondevila M., Freire A., Santos C., Carracedo A., Lareu M.V., Gusmão L. & C.M. Vullo. (2012). Typing short amplicon binary polymorphisms: Supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. 6 (4): 469-476.

Rosemberg N.A., Mahajan S., Ramachandran S., Zhao C., Pritchard J.K. & M.W. Feldman (2005). Clines, clusters, and the effect of study design on the inference of human population structure, *PloS Genetic* 1 (6):E70.

Rosado F.F. (2008). Análisis de 15 loci tipo short tandem repeats (STR) en la población de Paraguay para su uso en identificación humana. Tesis Doctoral, Universidad de Granada.

Smalldon K.W. & A.C Moffat. (1973). The calculation of discriminating power for a series of correlated attributes. *J Forensic Sci* 13: 291-295.

Saiki R.K., Gelfand D.H., Stoffel S., Scarf S.J., Higuchi R., Horn G.T., Mullis K.B. & H.A Erlich. (1988). Primer-Directed Enzymatic of DNA with a thermostable DNA polymerase, *Science*: 239(4839):487-91

Saiz M. (2014). Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population. *Forensic Science International: Genetics* 9 e27–e29.

Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M. & L.D Stein. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*: 409(6822):928-33.

Santos N.P., Ribeiro-Rodrigues E.M., Ribeiro-Dos-Santos A.K., Pereira R., Gusmão L. Amorim A., Guerreiro J.F., Zago M. A., Matte C., Hutz M.H. & S.E

Santos. (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat.*31:184-190.

Slatkin M. (1991). Inbreeding coefficients and coalescence times. *Genet. Res. Camb.* 58: 167-175.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Slatkin M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet.*9:477–85.

Tishkoff S.A., Reed F.A., Friedlaender F.R., Ehret C., Ranciaro A., Froment A., Hirbo J.B., Awomoyi A.A., Bodo J.M., Doumbo O., Ibrahim M., Juma A.T., Kotze M.J., Lema G. Moore J.H., Mortensen H., Nyambo T.B., Omar S.A., Powell K., Pretorius G.S., Smith M.W., Thera M.A., Wambebe C., Weber J.L. & S.M. Williams. (2009). The genetic structure and history of Africans and African Americans, *Science* 324, 1035-1044.

Thompson E. & S. Guo. (1992). Performing the Exact Test of Hardy-Weingberg Proportion of Multiple Alleles. *Biometrics*, Vol .48 (2): 361-372.

Watkins W.S., Ricker C.E., Bamshad M.J., Carroll M.L., Nguyen S.V., Batzer M.A., Harpending H.C., Rogers A.R. & L.B. Jorde. (2001). Patterns of ancestral human diversity: An analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* 68: 738–752.

Weber J.L. & K.W. Broman. (2001). Genotyping for human whole-genome scans: past, present, and future. *Adv Genet* 42: 77–96.

Weber J.L., David D., Heil J., Fan Y., Zhao C. & G. Marth. (2002). Human Diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.* 71: 854-862.

Wenk R.E. (2004). Testing for parentage and kinship. *Curr Opin Hematol.* 5:357-61. PMID- 15666661.

Wicks S.R., Yeh RT., Gish WR., Waterston RH. & R.H.A. Plasterk. (2001). Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet.* 28:160-164.

Warren S.T., Zhang F., Licameli G.R. & J.F Peters. (1987). The fragile X sites in somatic cell hybrids: An approach for molecular cloning of fragile sites. *Science* 237: 420–423

Yang N., Li H., Criswell L.A., Gregersen P.K., Alarcon-Riquelme M.E., Kittles R., Shigeta R., Silva G., Patel P.I., Belmont J.W. & M.F Seldin. (2005). Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet.* 118(3-4):382-392.

Zaumsegel D., Rothschild M.A. & P.M Schneider. (2013). A 21 marker insertion deletion polymorphism panel to study. *Forensic Science International: Genetics.* 7(2):305-312.

